INTO THE UNKNOWN:

EXPLORATION OF SEARCH ENGINES' RESPONSES TO USERS WITH DEPRESSION AND ANXIETY

by

Ashlee Milton

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Computer Science

Boise State University

August 2021

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Ashlee Milton

Thesis Title: Into the Unknown: Exploration of Search Engines' Responses to Users with Depression and Anxiety

Date of Final Oral Examination: 24 June 2021

The following individuals read and discussed the thesis submitted by student Ashlee Milton, and they evaluated the presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Maria Soledad Pera, Ph.D.	Chair, Supervisory Committee
Michael Ekstrand, Ph.D.	Member, Supervisory Committee
Francesca Spezzano, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Maria Soledad Pera, Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

ACKNOWLEDGMENTS

The writing of this thesis happened during one of the most turbulent times in my life which makes me extremely thankful for all the support and guidance I received.

First and foremost, I would like to thank my mentor and advisor Doctor Maria Soledad Pera, who saw my potential as an undergraduate and pushed me to be the researcher I am today. While there was more than a few late nights, tight deadlines, and mental breakdowns you stuck with me and made sure I did not loss sight of the goal or myself, and for that I am truly thankful.

I would like to acknowledge my committee, who were extremely flexible and always willing to answer any questions I had. Additionally, I would like to thank my colleagues from the People and Information Research Team (PIReT) who were always willing to help, whether it be debugging code or interpreting my hand gestures as words I could not remember.

Finally, I would like to thank my friends and family. While most of you still do not get what I do or why I am still in school, you always listen to me ramble about research anyways.

ABSTRACT

Mental health disorders (MHD) are a rising, yet stigmatized, topic. With statistics reporting that one in five adults in the United States will be afflicted by a MHD in their lifetime, researchers have begun exploring the behavioral nuances that emerge from interactions of these individuals with persuasive technologies, mainly social media. Yet, there is a gap in the analysis pertaining to a persuasive technology that is part of their everyday lives: search engines (SE). Each day, users with MHD embark on information seeking journeys using SE. Every step of the search process for better or worse has the potential to influence a searcher's state of mind. In this thesis work, we empirically investigate what subliminal stimulus SE present to these vulnerable individuals during their searches. We do so by utilizing an information retrieval foundation that leverages data and techniques from psychology, social media, and natural language processing. Outcomes from this work showcase open problems related to query suggestions, search engine result pages, and ranking, that the information retrieval community needs to address so that SE can better support individuals with MHD.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

– **ISP** Information seeking process

– **MHD** Mental health disorder

– **Q** Query

– **QS** Query suggestions

– **RR** Retrieved resource

– **SE** search engine

– **SERP** Search engine result pages

CHAPTER 1

INTRODUCTION

Technology has become integrated into most facets of our lives; the way we interact with it has changed, and not always for the better [27, 2]. Persuasive technologies are designed to influence the behaviors or attitudes of individuals [28], but not everyone is swayed by technology in the same manner. Think about people suffering from mental health disorders (**MHD**), such as schizophrenia or bipolar disorder. Their rational decision-making is inhibited due to their emotional states [75], causing them to be more susceptible to persuasion than individuals not ailed with a MHD. This leads us to think that people ailed with MHD would interact with and be affected by persuasive technologies differently. With mental illness being a persistent issue [3], one now openly discussed both on- and off-line [51], it is imperative to identify and understand the consequences that unknowingly occur when people suffering from MHD engage with persuasive technologies.

Search engines (**SE**) are a ubiquitous persuasive technology, yet there is little information regarding how MHD users engage with it. Consider Figure 1.1, a snapshot of Google's snippets for the query "waste of space", a phrase a person with depression may say or think. Among the resources retrieved we find dictionaries, which on the surface are benign. However, upon closer examination of these resources we see phrases like "worthless person", "He's a complete waste of space", "fat bastard", and "I'm just a waste of space". It would not be surprising for someone battling a MHD to feel distressed by

Figure 1.1: Google's SERP for the query "waste of space."

such phrases. As reported by the National Institute of Mental Health, one in five adults in the United States suffers from a MHD [55]. With millions of individuals turning to SE regularly [1, 21], searchers suffering from MHD are a large population. This makes it crucial to investigate how this diverse user group interacts with and is potentially impacted by SE, as that will reveal knowledge gaps that the information retrieval community must address so that SE can better respond to users with MHD.

Scrutinizing SE behaviors when responding to MHD users is a complex issue, as individuals with MHD are particularly sensitive to both internal and external emotions [5]. To set the foundation for understanding this concern, we focus on an *algorithmic perspective*

as a starting point. We conduct an empirical exploration of SE functionality using a **subliminal stimuli** lens, where subliminal stimuli is an amalgamation of both affect[1] and MHD indicators. Specifically, we examine the subliminal stimuli present in text during each stage of an information seeking process (**ISP**), as defined by Kuhlthau [38], where SE directly interact with users. The ISP consists of six stages: initiation (lack of knowledge), selection (topic identification), exploration (gathering information), formulation (evaluating information), collection (information found), and presentation (completing the search). By examining all stages that prompt SE response, we form a well-rounded view of this process. We map each ISP stage to a specific SE functionality: (i) query suggestions (**QS**)–selection, (ii) search engine results pages (**SERP**)–exploration and formulation, and (iii) retrieved resources (**RR**)–collection and presentation. Moreover, we consider the affect and MHD indicators present in users **queries**–initiation, to characterize the users who prompt these algorithmic responses. As a counterpoint to help us recognize whether subliminal stimuli from SE is biased for the population under study, or just the result of typical algorithm behavior, we consider the affect and MHD indicators present in queries and subsequent SE responses to traditional searchers, which we treat as our **control** group.

To manage scope, we center our study on *English* speaking *adults*, as a large portion suffer from a MHD (46.6 million in 2017 [55]), and popular commercial SE (Google and Bing), as they are mainstream among English speakers. Further, it is well-documented that the degree to which subliminal language affects users with MHD depends on the kind of MHD they have [46, 57, 45, 7]. To not overgeneralize, we only examine *depression* and *anxiety*[2]. With this thesis, we aim to answer the overarching question: *How do SE respond on a subliminal level to users with MHD during their information seeking journey?*. To

---

[1]We use the psychology view of *affect*: "any experience of feeling or emotion" [6].
[2]From here on, whenever we state MHD, we refer to depression and anxiety.

guide our exploration, we define two research questions:

1. What subliminal stimulus do SE responses project directly onto users with MHD?

2. How do subliminal stimulus of SE responses indirectly change through the ISP for MHD users?

For analysis purposes, **text samples** (i.e., query suggestions, snippets, web resources) that capture ISP interactions that represent MHD searchers, as well as the control group are vital. Unfortunately, large scale query logs from popular SE are rarely accessible for research. To further complicate the issue of data, interactions with SE from people with MHD are not available. For these reasons, we allocated research efforts to build synthetic datasets. To do so, we turn to Reddit and Yahoo Webscope. These data sources are used to build the datasets imitating interactions of MHD users (and the control group) with SE. We then examine these datasets to gauge SE reactions to traditional and MHD searchers using lexicon- and machine learning-based techniques [61, 52, 44].

The contributions of this work include: (i) generating *subliminal profiles* of SE responding to interactions with MHD users, (ii) creating three domain specific lexicons, one based on social media posts for anxiety and two using psychological surveys for depression and anxiety, as well as (iii) an in-depth look at the current state of SE responses to MHD searchers through the inspection of MHD and control users interactions, highlighting limitations of popular SE and implications informing future research. Our exploration could be used as a framework for future research on this subject from different perspectives, expanding the knowledge of the subliminal stimulus users with MHD face when interacting with SE. Outcomes serve as a foundation to inform SE design to support individuals affected by diverse MHD. Implications on this work could extend into a variety of areas, for example examining whether lessons learned from our work apply internationally, as MHD

present themselves and are dealt with differently among varying languages and cultures. Mental illnesses are also prominent among younger populations, particularly teens with depression and anxiety. Therefore, our investigation can help set the stage for further inquiries into the problems when focused on this audience. Human-computer interaction can also benefit from our work, i.e., identified trends could be leveraged into the design of SERP that avoid displaying information that may exacerbate the symptoms of some MHD. Moreover, differences in how subliminal language is used by MHD versus mainstream users could inform the design of adaptive applications that can aid users who are seeking or receiving treatment for MHD.

The rest of this manuscript is organized as follows: We start with Chapter 2, where we provide the thesis statement. In Chapter 3, we discuss background and related literature. In Chapter 4, we detail the data that we use in our exploration; we also offer descriptions of the strategies we employ to represent the subliminal stimuli of our data. In Chapter 5, we detail the experiments conducted in order to answer our research questions and provide a discussion of our findings. Lastly, in Chapter 6, we offer concluding remarks, address limitations of our exploration, and present future research directions informed by our findings.

CHAPTER 2

THESIS STATEMENT

We aim to explore how search engines respond to users suffering from depression and/or anxiety during each stage of the information seeking process. We examine the content presented via query suggestions, search engine result pages, and retrieved resources in response to queries representing searchers with depression and/or anxiety. To do so, we depend upon theoretical foundations from information retrieval, in addition to techniques from psychology, social media, machine learning, and natural language processing, which will enable comparison of subliminal stimulus induced by search engines responding to depressed/anxious users versus traditional users for contextualization.

CHAPTER 3

RELATED WORK

Mental health disorders are a prevalent concern, one that has received attention from researchers and practitioners, as evidenced by the many web and mobile applications that have been developed as a way to track and treat symptoms [36, 58]. These programs and associated research range from using mobile phones and wearable technology for depression tracking [76] to mobile software for depression assessment [20]. The common denominator among these applications is that they are based on how people with MHD turn to technology but not how technology influences them, which is the focal point of our research. In the rest of this chapter, we discuss the background and related literature that contextualize this thesis work.

3.1 MHD and Social Media

From a persuasive technology perspective, MHD literature is focused on the social media domain. Mainly, the ability to identify users with depression from social media posts [23, 72, 62, 24, 59]. Depressed users have not been the only ones considered, as researchers have also studied the linguistic qualities of social media posts by users with other MHD, such as schizophrenia [16]. Findings from the research conducted thus far are the result of examining the text of social media posts (primarily the vocabulary, syntax, and linguistic style of posts), the interactions made by MHD individuals with the platforms themselves

(e.g., number of posts and retweets) [74, 66], as well as the trends in interactions of MHD users with other social media users [74]. In an attempt to detect the state of mind of social media users, several researchers have also considered trends and linguistic styles of users with MHD. An in-depth overview of current research efforts allocated to achieve this goal can be found in the recent survey by Ríssola et al. [60], where the authors summarize the many computation methods that have been developed to detect a social media user's state of mind. While there has been progress in understanding MHD in social media, our current work is focused on SE.

3.2   MHD and SE Interactions

Research exploring the relationship between MHD and SE is in its infancy. From a user perspective, Campbell et al. [18] discuss help-seeking behavior of users with MHD, i.e., searchers looking for resources to understand and help with MHD. Zhu et al. [83], on the other hand, use query logs from a university webserver to predict users suffering from depression. Similarly, Zaman et al. [80] identify searchers with self-esteem issues from user-provided Google search histories. Most recently, Birnbaum et al. [16] have contemplated the feasibility of detecting the early onset of MHD from query logs. Instead, Xu et al. [77] turn to query logs to evaluate the degree to which mood influences users' interactions with SE. While not focused on MHD, Moshfeghi and Jose [50] bring up an interesting point often overlooked when scrutinizing query logs for MHD-related tasks: query logs capture user interactions, but do not provide specific search tasks. This is a limitation, as not all search tasks are the same and depending on the task or users' intentions, different emotions can be experienced at varying levels. Ever since March of 2020, the world has been in the midst of a global pandemic (i.e., COVID-19). This has

prompted researchers to study if and how search trends for mental health have changed [35, 8]. Outcomes revealk that mental health queries are more prevalent now, evidencing the needs for explorations such as the one we present in this thesis.

In short, previous work has investigated user interactions from the perspective of help-seeking, self-esteem, and mood. Unfortunately, none of these research contributions study the search systems themselves in order to shine a light on the potential that SE responses have to alter the mental well-being of users with MHD.

3.3 MHD and the Information Seeking Process

Interactions of both traditional and non-traditional users with SE at different ISP stages have been widely explored. Representative research works include those by Chelaru et al. [19] who investigate the sentiment present in queries but does not consider emotions. Azpiazu et al. [9] and Locke et al. [43] respond to the QS needs of children and domain experts, respectively. However, there is a gap in QS research related to users with MHD. As for SERP, the work of Zhang et al. [82] utilize visual aspects of SERP to estimate the relevance of a resource, whereas Ling et al. [41] use ensemble models to predict ad click-through rates on SERP. The works of Gossen [29] and Morris et al. [49] center on children and dyslexic persons' experiences with SERP. Regardless of the ISP stage, we note that there is a lack of literature pertaining to affectitve analysis and representation of MHD users. Few existing initiatives in this area include the work by Till et al. [70], who investigate the differences that have appeared in web page contents related to the topic of suicide over the last five years but do not consider the affect expressed in the resources. Kazai et al. [37] and Demartini and Siersdorfer [25] investigate the emotions, sentiments, and opinions emerging from web resources, yet only in response to queries formulated by

traditional users. Additionally, Landoni et al. [39] explore SERP and emotions, but in their case the population under study are children.

While the works investigate SE functionality or affective responses at each ISP stage, none directly study the comprehensive subliminal stimulus responses of different SE functionality for MHD individuals. In our study, we take a first step towards understanding the gaps we see in the literature regarding the information seeking journey of MHD searchers, to determine what stimuli are being pushed onto users with MHD who are already struggling with their own emotions.

CHAPTER 4

EXPERIMENTAL SETUP

In this chapter, we discuss the experimental setup of the empirical study conducted in order to answer our research questions.

4.1 Data Collection

Query logs from mainstream SE are seldom available for research and, to the best of our knowledge, non-existent when specifically capturing interactions initiated by MHD users. Consequently, to enable exploration of subliminal stimulus when SE respond to MHD users, we first need to gather data that is representative of MHD searchers' interactions with SE. We start with emulating queries that MHD users would formulate using the the research presented by De Choudhury et al. [24]. We treat as queries phrases we extract from Reddit posts (4,418 synthetic queries: 1,200 unigrams and the remaining n-grams). Reddit offers several subReddits for people with MHD so the subReddit's posts capture the language and topics used by MHD users in an online forum environment.

We use the collected queries to mimic users interactions with SE and elicit responses which result in text samples from QS, SERP, and RR for both Google[1] and Bing[2] using their respective API's. When collecting data, we only record the first SERP, as users do not

---

[1]https://developers.google.com/custom-search
[2]https://www.microsoft.com/en-us/bing/apis/bing-web-search-api

often go past the first page when looking at search results [64]. For each SERP result, we extract the title, snippet, as well as full web content from corresponding web resources.

It is important to note that, we do not have access to user-system interactions, i.e., click-through data, which is why we exclude this information from our analysis. Moreover, we focus our analysis on interactions assumed to be initiated by English speaking users, therefore we only consider queries in English.

Additionally, we also consider a control group, comprised of traditional users, in our analysis and thus, need data representing their interactions as well. We use a sample of queries made available for research purposes by Yahoo Webscope's [79] (4,458 queries: 1,211 unigrams and the remaining n-grams) to represent our control group. We follow the aforementioned procedure to gather the appropriate text samples.

4.2 Establishing Stimulus Vectors

To get a full view of the subliminal stimulus expressed by SE, we explore the text samples collected in Section 4.1 (e.g., query suggestions, resource titles, snippets, and web page content) of both MHD users and our control from different perspectives: (i) intensity of affect, (ii) prominence of MHD terminology, and (iii) evidence of MHD. For each text sample, we create a *stimulus vector*, which accounts for each of the aforementioned perspectives and is illustrated in Figure 4.1. This stimulus vector serves as a representation of the affect and MHD indicators present in the corresponding text sample.

4.2.1 Intensity of Affect

Lexicons are a starting point for discerning affective language in text samples. The ones most relevant to our study are those specific to sentiment and emotion. Following the
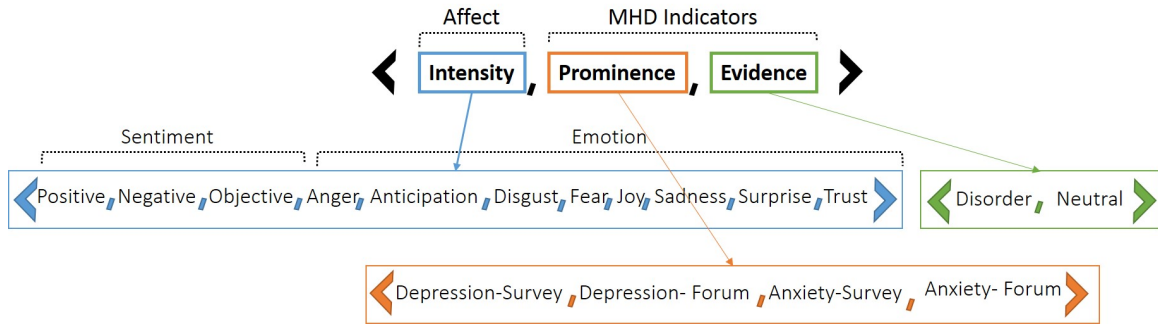
Figure 4.1: Structure of the stimulus vector representing each text sample

framework presented by Kazai et al. [37], which represents the sentiments and emotions emerging from data as a distribution of intensity scores, we create an affect **intensity** vector depicting the affect distribution of text samples. For **sentiment**, we use SentiWordNet [10]. In this lexicon, sentiment is represented as `Positive`, `Negative`, and `Objective`. For **emotion** identification, we use Emotion Intensity Lexicon (NRC-EIL) [61], a lexicon that represents words as vectors of `Anger`, `Anticipation`, `Disgust`, `Fear`, `Joy`, `Sadness`, `Surprise`, and `Trust`. Both lexicons represent the affects of a word on a scale of 0 to 100, with 100 indicating a word is evocative of a given affect. To produce the affect Intensity vector of a text sample, we average the affects for each word in the sample over its length.

4.2.2 Prominence of MHD

The terminology that MHD individuals both use and respond to is noticeably different than that of traditional users [14]. Thus we also investigate the prominence of terms commonly associated with MHD in text samples, i.e., the frequency of domain-specific terminology in text samples. For depression-related terms, we combine the lexicons made available by Losada and Gamallo [44] into a single one referd to as `Depression-Forum` consisting of 899 terms.

To our knowledge, there are no domain-specific lexicons for one of the MHD under study, anxiety. Consequently, we adopt the procedure outlined by De Choudhury et al. [23], which relies on the pointwise mutual information and log likelihood ratio of bigrams (generated with a regex) on Yahoo Answer! posts related to topics of mental health, to build a lexicon for depression. In our case, we use the Reddit data collected by the authors of [65] over a three month period in 2017 from the subReddits: r/Anxiety, r/SocialAnxiety, and r/PanicParty, for lexicon generation. To remove frequent terms (based on TF-IDF) that may overlap with those identified to be part of the lexicon, we use a subset of 1.6 million Wikipedia articles, unlike the full Wikipedia in [23]. Further, as we are interested in finding terms related to anxiety, we use the regex "anx*" in our bigram generation. This results in our `Anxiety-Forum` lexicon, which contains 79 terms.

We are aware that as our depression and anxiety lexicons are constructed from social media posts, they may overlook formal terms that psychologists would consider in MHD diagnosis. Therefore, we construct two new lexicons, one for depression and another for anxiety, comprised of terminology that we infer from psychological assessments:

1. State Worry Questionnaire (PSWQ) [68]

2. Liebowitz Social Anxiety Scale (LSAS-SR) [40]

3. Hamilton Anxiety Rating Scale (HAM-A) [31]

4. Anxiety Symptoms Questionnaire (ASQ) [11]

5. Generalized Anxiety Disorder (GAD) Screening Tool [54]

6. Beck's Depression Inventory [12]

7. Patient Health Questionnaire-9 (PHQ-9) [67]

8. Hamilton Depression Rating Scale (HAM-D) [32]

9. Montgomery and Åsberg (MADRS) Depression Rating Scale [48]

10. EQ-5D [34]

11. Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [5]

These assessments are a variety of diagnostic tools used by medical professionals, e.g., doctors, counselors, or psychologists, to determine if a patient is suffering from depression and/or anxiety. Each assessment is comprised of statements to which a patient will respond using a scale, usually either how much they agree or how frequent they experience a statement. As these statements contain language meant to resonate with people with MHD, we depend upon an impartial assessor to identify and extract most frequent keywords in the assessments related to each MHD. This yields two lexicons: `Depression-Survey` containing 47 terms and `Anxiety-Survey` comprised of 17 terms.

Using the aforementioned lexicons, we generate for each text sample a **Prominence** vector, comprised of the four scores, one for each lexicon. Each score is a proportion of the total number of words in the corresponding lexicon that are in a text sample over the total number of words (including stop words) in the sample. The denominator acts as a normalization factor to ensure that text sample length does not influence score computation.

4.2.3 Evidence of MHD

Contemplate the statement "I will never be happy again". From the individual keywords in the statement, one could assess it to be "happy" in tone; associating the word "never" to other terms in the phrase reveals the real tone of the phrase: sadness. Further, "I" is prefacing a negative emotion; a linguistic style commonly seen with depressed individuals,

who are known to frequently use self-referencing statements with negative emotions. When examining terms in isolation, it is possible to miss the nuances that could be inferred from text as a whole. With this in mind, we explore text samples from a holistic standpoint. For this purpose, we adopt the mental health multi-class classification strategy introduced by Murarka et al. [52]. This strategy utilizes a RoBERTa model [42] that explores a text (specifically title and body of Reddit posts) as a whole in order to determine the likelihood of said text conveying the writing patterns of individuals affected by anxiety, depression, ADHD, PTSD, or bipolar.

We adapt the strategy in [52] to act as a binary classifier, as we are only interested in determining if a SE response is indicative of MHD, as per our definition. Additionally, we alter the manner in which text is cleaned as, unlike the original strategy, we also remove special characters, expand contractions, and correct misspelled words. We train the adapted strategy using the same libraries, parameters, and data as in [52], with text samples truncated to 512 tokens for 10 epochs using an Adam optimizer with a learning rate of 0.00001 and a dropout layer with a 0.3 probability implemented with the libraries PyTorch and Huggingface.[3]

We use our trained model to create an MHD **Evidence** vector for each of our text samples. This vector captures the probability scores for each class, `Disorder` or `Neutral`, in the range of 0 to 100, with 100 denoting the text is indicative of a respective class.

4.3 Generating Subliminal Stimulus Profiles of SE

The stimulus vectors of text samples provide insights into the individual samples but not SE responses in general, which is the goal of our study. With this in mind, we combine

---

[3]We empirically verified that the model yields a 97% accuracy for classification, which is why we deem it applicable for our task.

the aforementioned vectors for a given ISP into a single *subliminal stimulus profile* which serves as a snapshot that captures the subliminal stimulus (i.e., affect and MHD indicators) presented to SE users as a result of their interactions with the system.

We first create an *overall profile* by aggregating all the text sample vectors that correspond to an ISP stage. Aware that text samples can appear more than once when SE respond to users, we want to combine these sample vectors in a way that will mitigate any bias that could be introduced in an overall profile. We also build a *by-query profile*, which first groups text sample vectors by the query that initiates their generation and averages them; this is followed by aggregating the per query vectors. The overall and by-query profiles provide a combined view of SE functionally, but we are also interested in how rank plays a role in subliminal stimuli. We know that users pay attention to the order in which information is presented and this order can influence users' interactions. This leads us to create a *rank-1 profile*. In this case, we average the first ranked text sample vector for each query that initiates its generation.

Example. To illustrate the generation of each profile, suppose we have a data containing the text samples for the queries "apple" and "orange" and we are considering SE functionality QS. For an overall profile, we would take every QS text sample vector produced in response to "apple" and "orange" and average them altogether resulting in the profile for QS. To create the query-based profile, we would first average all the QS sample vectors for just "apple", then do the same for "orange", which would result in two per query vectors. These two vectors would then be averaged together generating the query-based profile for QS. In our example query log, let's assume each query has three QS associated with it. When constructing the rank-based profile for QS, we would take the first ranked QS sample vector responding to both "apple" and "orange" and average them, generating the rank-1 profile for QS.

4.4   Considering External Variables

Recall that, in Chapter 1, we stated that the aim of the study described in this thesis was to understand the subliminal stimulus that SE present to users with MHD at different stages of the ISP. To enable this investigation, we collected data over an extended period of time—from December 2019 to May 2021—that we use to simulate the synthetic user-interactions with SE across the ISP, along with the corresponding SE-responses which we study. Due to the dynamic nature of the web, the responses SE present regarding a query one day, may not be the same as the next, whether it be the change of a result or the shifting in affect of a snippet. Thus, it is possible that the affect SE are displaying in response to MHD users information needs can shifts over *time*. Moreover, it cannot be denied that the last two years have been particularly turbulent. There have been several *major events* that have occurred, causing prolonged stress on users, and in some cases triggered the development of MHD [69, 22]. Naturally, these events have also inundated SE with new resources about said events and, given the impact on users, these resources may have wildly different affects then other resources SE present.

We posit that the aforementioned external factors could impact data collection and indirectly skew the analysis of SE responses. Consequently, we are obliged to consider if and how external factors alter the stimulus that would otherwise be portrayed in SE responses. In the rest of this section, we discuss the findings of several experiments we conducted in order to gather insights into the impact external factors can have our collected data (described in Section 4.1). To control scope, in both cases we examine the overall stimulus captured on SERP responses generated by Google.

The influence of time. Given a query, due to its adaptive nature, a SE might include different resources on its SERP over time. This can cause the stimuli on SERP to alter.

Thus, we use a representative sample of queries (690 queries uniformly distributed across MHD users and our control) from Section 4.1 and generate SERP at different *times of the day* (12am, 6am, 12pm, and 6pm) and different *days of the week* (Sunday, Monday, Wednesday, Friday, and Saturday). We then compute the vectors detailed in Section 4.2 and generate the overall stimuli profile, as outlined in Section 4.3, of both MHD users and our control. For adjacent pairs of profiles, e.g., Sunday and Monday, Monday and Wednesday, 12am to 6am, or 6am to 12pm, we compare the stimulus expressed to denote any changes. Significant differences are reported across pairs of profiles based on a two tailed t-test ($p < 0.01$).

Table 4.1: Results of the pairwise comparison between intensity vectors generated using data collected on different days of the week and time of the day.

| Profile | Intensity | | | | | | | | | | |
| | *Sentiment* | | | *Emotion* | | | | | | | |
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Sunday	3.86	2.85	93.29	3.58	11.09	2.49	6.30	13.76	4.47	3.35	19.72
Monday	3.87	2.87	93.26	3.58	11.00	2.52	6.25	13.82	4.49	3.32	19.6
Wednesday	3.90	2.86	93.24	3.53	11.19	2.44	6.19	13.81	4.43	3.33	19.55
Friday	3.89	2.85	93.26	3.47	11.12	2.42	6.22	13.85	4.43	3.40	19.68
Saturday	3.88	2.84	93.28	3.52	11.14	2.45	6.17	13.86	4.43	3.39	19.64
12 am	3.92	2.86	93.22	3.53	11.18	2.52	6.31	13.79	4.48	3.32	19.58
6 am	3.89	2.86	93.24	3.52	11.17	2.43	6.19	13.80	4.46	3.35	19.60
12 pm	3.90	2.86	93.24	3.53	11.19	2.44	6.19	13.81	4.43	3.33	19.55
6 pm	3.87	2.87	93.26	3.54	11.18	2.45	6.23	13.72	4.50	3.36	19.63

As captured in Table 4.1, it emerges from the rows corresponding to days of the week we see that `Positive` increases from Sunday until Wednesday then decline through the remaining of the week. From the time of day rows, `Joy` tends to increase until 12pm, then decreasing through the rest of the day; we also observe decreases in `Sadness` up to 12pm which then increases until the end of the day. All together, even though there are some slight changes in stimulus, none of the differences across days of the week and time of the day considered for data collection to generate the stimuli profiles reported in Table 4.1 are

statistically significant.

The impact of major events. With the impact events can have on users, SE responses may all be expressing altered stimuli on SERP. Consider the two main major events to happen in the United States in 2020: COVID and the presidential election. These event have been extremely emotionally devastating for some users. Thus we utilize the same sample of queries used in our time experiment and generate SERP in December 2019 prior to COVID or the election, and October 2020 with the election at its peak and COVID having been a reality for over half a year. Per the procedures described in Section 4.2, we generate vectors and use them to create overall stimulus profiles as outlined in Section 4.3 for each set of retrieved SERP. To detect significance differences between profiles, we use a two tail t-test.

As captured in Table 4.2, the subliminal stimuli has shifted significantly across profiles. There are increases from the data collected in 2019 to data in 2020 in `Positive`, `Negative`, and `Objective`, while all other affects have decreased. To dig further into the shifts we observe, as two separate events happened over the same time period, we split the SERP gathered in 2020 into two new profiles: one made using results related to COVID and the other the presidential election. We compare the overall subliminal stimulus profiles to observe if the events have individually altered the subliminal stimulus profiles of SE responses.

In the last two rows of Table 4.2, are the results of the generated profiles for COVID and the election. When looking at sentiment, both profiles have increases in `Positive` and `Negative` with respect to the 2019 profile, but the COVID profile has a bigger increase to `Negative`. Similar trends can be seen in the other affects, with `Anticipation` and `Joy` both decreasing for COVID and the election but, COVID decreases by a larger margin. The COVID profile does not have significant changes in `Anger`, `Fear`, and `Sadness`, while the

Table 4.2: Intensity vectors computed on data collected before January 2020 and after October 2020; blue denotes vector components that significantly different from their counterpart computed on data collected prior to January 2020 ($p < 0.05$).

Profile	Intensity										
	Sentiment			*Emotion*							
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Before-All	3.83	2.9	93.27	3.75	11.4	2.64	6.27	13.76	4.74	3.44	18.49
After-All	5.88	3.84	90.28	0.89	2.59	0.64	1.72	3.43	1.25	0.78	4.28
After-COVID	4.87	4.6	90.53	1.71	2.38	0.7	5.68	2.05	3.66	1.08	4.66
After-Election	4.54	3.71	91.74	1.41	1.98	0.84	1.36	1.58	0.67	2.59	4.87

election only doe not have a significant change in `Surprise`. It is clear that each event had different impacts on SE responses, but neither fully encompass the changes seen between the SERP collected in 2019 and 2020.

Implications. Outcomes from both experiments reveal that while the external variable of time does not result in stimulus profiles that significantly differ, major events do causes shifts in the stimulus SE present to users. While the timeline for data collection in Section 4.1 should not result in altered SE responses, the aftermath of the most recent major events continues to cause impacts to the stimuli SE display to users. For these reasons, the analysis and discussion we report in Chapter 5 is based on data collected after October 2020, as it is more representative of the current state of SE responses and the subliminal stimuli they express.

4.5 Generating Datasets and Associated Subliminal Stimulus Profiles

Informed by the in-depth analysis reported in Section 4.4, we exclude from the data collected in Section 4.1 any text samples retrieved before November 2020. We use the remaining samples to construct the datasets and the corresponding stimulus profiles that are key to the analysis we conduct to understand how SE respond to MHD searchers (Chapter 5).

Each dataset encompasses the responses a SE presents to a particular user group at a given stage of the ISP. With this in mind, we generate datasets that account for all possible combinations of user group, SE, and and ISP stage: **{MHD (M), Control (C)}–{Google (G), Bing (B)}–{Q, QS, SERP, RR}**. For example, **M-G-QS** is the dataset containing text samples from Google's QS resulting from synthetic queries emulating those belonging to MHD searchers.

From each of the resulting datasets we create the respective subliminal stimulus profiles, as described in 4.3. In naming these profiles, we use the same naming convention we used for the datasets: **{M, C}–{G, B}–{Q, QS, SERP, RR}–{Overall (O), By-Query (BQ), Rank-1 (R1)}**. For example, **M-G-QS-O** refers to the stimulus profile obtained using dataset **M-G-QS**, i.e., the stimulus inferred from query suggestions generated by Google in response to (synthetic) queries formulated by MHD searchers.

4.6   Testing for Significance

To understand the stimulus SE convey to users with MHD, we compare and contrast in Chapter 5 the various pairs of subliminal profiles generated in Section 4.5, i.e., M-G-QS-O vs. C-G-QS-O, M-B-SERP-R1 vs M-B-RR-R1, etc. We use a two tailed t-test with a Bonferroni correction (with $\alpha = 0.01$ and $\alpha = 0.05$ and the number of tests $N = 15$, which is the number of vector component) to indicate changes in stimulus across pairs of profiles that are statistically significant.

CHAPTER 5

EXPERIMENTS AND ANALYSIS

In this chapter, we describe the in-depth analysis and findings that result from the experiments we conducted in order to understand, from different perspectives, the subliminal stimuli that SE present to users with MHD in response to their information needs. We explore SE responses at each of the main stages of the ISP; we also consider the stability in stimuli, or lack thereof, across different stages of the information seeking journey, e.g. queries to query suggestions. Along the way, we compare the subliminal stimulus of SE responses to MHD searchers with those presented to the control group, as that can offer further insights into the affect and MHD indicators MHD searchers are subjected to.

5.1 RQ1: What subliminal stimulus do SE responses project directly onto users with MHD?

Millions of users turn to SE every day to satisfy their information needs, including those living with MHD, who are known to be easily influenced by external stimuli [5]. In the case of this particular user group, it would then be plausible to think that exposure to any adverse stimulation from SE may alter their states of mind. This is what led us to explore what subliminal stimuli SE directly convey to MHD searchers through query suggestions (**QS**), search engine result pages (**SERP**), and retrieved resources (**RR**). With this exploration, we aim to identify the subliminal stimuli exhibited by the profiles themselves and determine

whether the manner in which profiles are produced impact the observed stimuli. To do so, we turn to the profiles inferred from each of the datasets introduced in Section 4.5 (excluding query datasets), which we examine from diverse perspectives. We depict trends observed when aggregating stimulus profiles of text samples *overall* and *rank-1* on each ISP stage in Figure 5.1.

5.1.1 The Implicit Stimuli in Query Suggestions

As the first SE response that MHD users encounter in their information seeking journey, QS have the ability to change the direction of a search session. We start our exploration on QS by first dissecting Google's QS profiles in Table 5.1, i.e., M-G-QS-O, M-G-QS-BQ, and M-G-QS-R1. When considering the **Intensity** vector in all three profiles, `Negative` is always higher than `Positive`; `Anticipation`, `Fear`, `Joy`, `Sadness`, and `Trust` are higher than the remaining emotions in the *Emotion* vector. Even though we expected some of the high scores in the *Emotion* vector–by nature of the emotions often associated with MHD, like sadness and fear–two stimuli jump out as peculiar: `Joy` and `Trust`. To inspect what could cause these stimuli to be so high, we look at some QS for which their corresponding individual profiles also display high scores for the stimulus in question. We noted that these sample QS included terms like truth, love, and compassion (high `Trust`), as well as happiness, cheerful, and wonder (high `Joy`). While these terms can be used by individuals who have MHD, they are usually prefaced by negating words, like not or never, changing the context of the intended connotation of these phrases. For example, "cheerful" would appear very joyful, but "never cheerful" actually has quite a sad connotation. The unexpected high scores for `Trust` and `Joy`, coupled with the high `Negative` score, lead us to believe that the word independent assumption of the approach used to build **Intensity** vectors is the culprit for the high scores computed for upbeat emo-
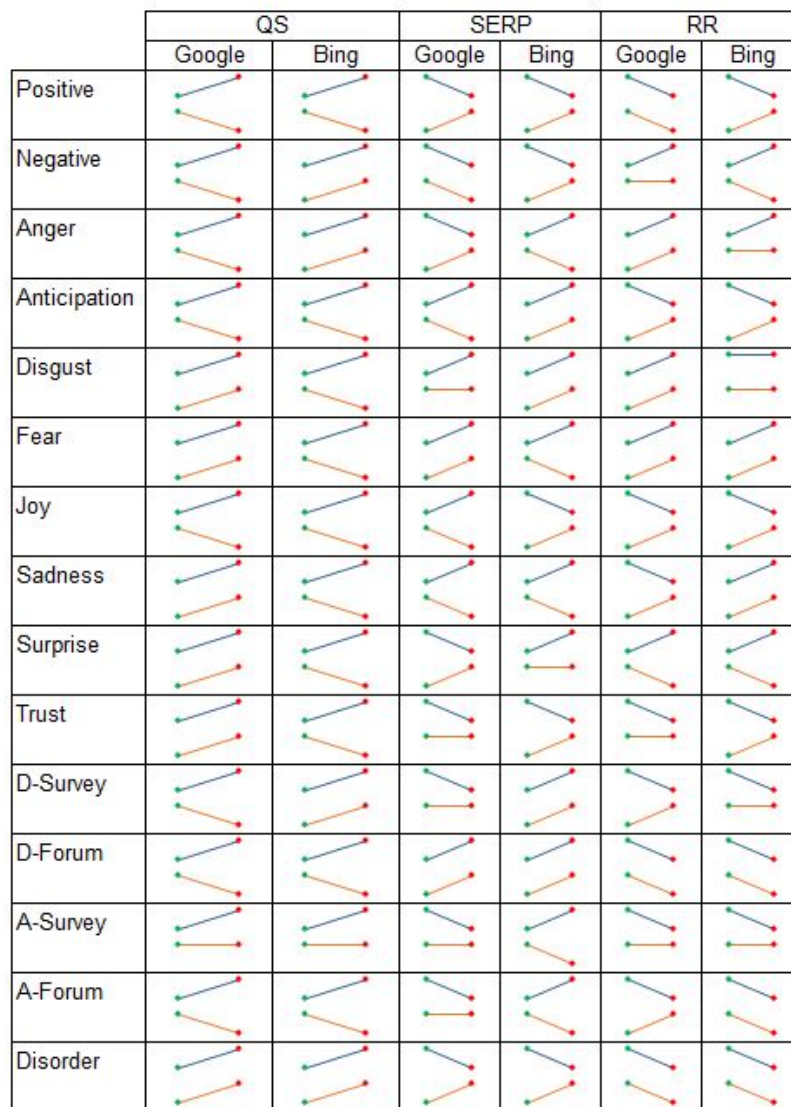
Figure 5.1: Subliminal stimulus trends between overall and rank-1 aggregation profiles for QS, SERP, and RR for Google and Bing. Trends lines in blue refer to MHD searchers, orange is control group. Green points refer to overall profiles, red is rank-1.

tions and therefore these scores do not necessarily represent the stimulus conveyed by the SE. On all three profiles `Depression-Forum` and `Anxiety-Forum` scores are noticeably higher than `Depression-Survey` and `Anxiety-Survey`. Regardless of the lexicon used, depression is the most prominent MHD. It is worth noting that the survey-based lexicons

have less terms than the forum-based ones and that depression lexicons are richer than their anxiety counterparts, which could explain higher overall scores in `Forum` than `Survey` and *Depression* than *Anxiety*. Surprisingly, when probing **Evidence** vectors in these three profiles, `Disorder` is lower than `Neutral`, hinting at writing patterns associated with MHD users not being prominent among QS.

Table 5.1: Subliminal stimuli profiles of QS generated by Google, blue indicates significant differences between MHD profiles and their corresponding control counterparts ($p < 0.01$); purple ($p < 0.05$).

Profile	Intensity											Prominence				Evidence	
	Sentiment						*Emotion*					*Depression*		*Anxiety*		Disorder	Neutral
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum		
M-G-QS-O	4.87	5.00	89.29	1.67	2.73	1.10	3.62	4.93	2.86	0.94	5.10	1.08	5.98	0.37	1.78	38.03	61.91
C-G-QS-O	2.63	2.53	90.40	0.60	1.56	0.47	1.12	2.34	0.76	0.51	2.48	0.10	2.18	0.02	0.15	15.81	84.60
M-G-QS-BQ	4.87	5.01	89.28	1.67	2.72	1.11	3.61	4.91	2.84	0.93	5.07	1.07	5.98	0.37	1.77	37.99	61.95
C-G-QS-BQ	2.64	2.54	90.39	0.61	1.56	0.48	1.12	2.34	0.76	0.51	2.48	0.10	2.17	0.02	0.15	15.83	84.57
M-G-QS-R1	5.13	5.58	89.16	1.96	3.05	1.27	4.47	5.52	3.57	1.14	5.95	1.43	6.56	0.47	2.31	44.63	55.09
C-G-QS-R1	2.45	2.41	94.95	0.58	1.48	0.49	1.21	2.32	0.80	0.55	2.49	0.09	1.94	0.02	0.14	16.35	83.81

Despite for the most part being alike, there are some peculiarities distinguishing profile score distributions across aggregation strategies. The profile capturing stimuli of the top QS generated for each query (i.e., M-G-QS-R1) portrays significantly higher scores for `Negative`, `Fear`, `Sadness`, `Trust`, `Disorder`, and `Neutral` when compared to M-G-QS-O as well as `Fear`, `Disorder`, and `Neutral` when contested with M-G-QS-BQ. Moreover, `Disorder` in M-G-QS-R1 is higher than its counterparts on the profiles generated using the two other aggregation emphasizing that among the QS presented to a searcher, the very first one is in fact the one tied more closely to MHD than any of the others.

To contextualize the observations we have made thus far, we consider the profiles of the control group, C-G-QS-O, C-G-QS-BQ, and C-G-QS-R1. When comparing the **Evidence** vectors of MHD profiles with their respective counterparts in control group profiles it becomes apparent that although `Disorder` is lower than `Neutral` in all three control profiles, when considered against `Disorder` and `Neutral` in MHD profiles, it becomes

clear that Google's QS for MHD searchers contain more MHD indicators. In fact, it is evident that stimulus scores are more prevalent, with the exception of `Objective` and `Neutral`, in profiles associated with MHD searchers, i.e., scores in the control profiles are statistically significantly lower than those of the corresponding MHD profile. Additionally, the **Intensity** vectors in the profiles of the control group come across as more stable in, i.e., there are less spikes in stimulus scores than those observed in counterpart MHD profiles (see Figure 5.2). Setting against each other the corresponding **Intensity** vectors in MHD and control group profiles enable us to spotlight that `Anger`, `Fear`, and `Sadness` are particularly high in QS Google produces in response to MHD users. Interestingly, all the differences in scores in the **Prominence** vector between MHD and control profiles, excluding `Depression-Forum`, exhibit at least a 10 fold increase for MHD users over control group users. Moreover, `Sadness` is 4 times higher for MHD users than our control. These last two trends evidence that there is a major difference in the terminology presented to MHD searchers. A key insight emerging from analysis of the profiles for the control group is that they have the same elevated scores for `Joy` and `Trust` we see in the MHD profiles but not in `Negative`, adding credence to our previous theory on the discrepancy due to the word-independence assumption of the approach for intensity estimation.

Table 5.2: Subliminal stimuli profiles of QS generated by Bing, blue indicates significant differences between MHD profiles and their corresponding control counterparts ($p < 0.01$).

Profile	Intensity											Prominence				Evidence	
	Sentiment			Emotion								Depression		Anxiety		Disorder	Neutral
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum		
M-B-QS-O	4.69	5.44	89.83	1.50	2.54	1.03	3.26	4.73	2.58	1.03	4.84	0.78	5.36	0.3	1.59	35.27	64.58
C-B-QS-O	2.37	2.73	94.84	0.53	1.47	0.41	1.03	2.22	0.65	0.45	2.37	0.1	2.02	0.02	0.15	13.97	86.40
M-B-QS-BQ	4.4	8.55	87.01	1.4	2.35	0.96	3.13	4.42	2.5	0.94	4.62	0.77	5.01	0.3	1.59	37.48	62.29
C-B-QS-BQ	2.3	4.47	93.19	0.53	1.37	0.38	0.97	2.16	0.6	0.42	2.2	0.11	1.97	0.02	0.15	15.34	84.99
M-B-QS-R1	4.9	9.35	85.76	1.75	2.67	1.18	4.04	5.03	3.35	1.07	5.44	1.06	6.11	0.42	2.4	42.5	57.09
C-B-QS-R1	2.28	4.5	93.19	0.54	1.28	0.37	1.0	2.06	0.61	0.41	2.26	0.14	1.91	0.02	0.11	16.78	83.42

We also examine the subliminal stimuli of QS produced by Bing by looking at the stimuli distribution in the profiles captured in Table 5.2. We observe that there is a sig-
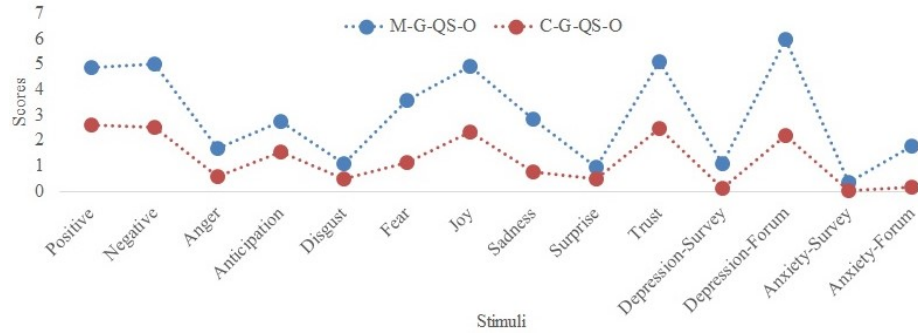
Figure 5.2: Representation of stimuli for M-G-QS-O and C-G-QS-O. For illustration purposes, we omit `Objective`, `Disorder`, and `Neutral` from the corresponding subliminal stimuli profile representation. It emerges from this image that while both profiles have similar distribution scores, there are visible spikes in the intensity of `Anticipation`, `Joy`, `Trust`, `Negative`, and `Depression-Forum`.

nificant decrease in the strength of `Objective` and `Neutral` when aggregating QS samples by-query (M-B-QS-BQ) as opposed to overall (M-B-QS-O), except for `Negative` and *Disorder* which significantly increase. Contrasting M-B-QS-R1 with M-B-QS-O and M-B-QS-R1 with M-B-QS-BQ, there is a significant decrease in `Neutral` and significant increases in `Fear`, `Sadness`, and `Disorder`. These findings indicate that the first ranked QS for Bing is the most stimulating QS compared to any other provided by the SE. To bring into perspective the diverse stimulus MHD users experience, we juxtapose the profiles of QS presented to MHD searchers versus those shown to traditional users. There is a proportionally large disparity between the stimulus scores in the **Intensity** vector in the control group profiles and those on MHD profiles. Similarly, `Disorder` scores are lower in control group profiles, suggesting that MHD writing styles are not necessarily common among QS presented to traditional searchers. We also note that regardless of the aggregation strategy there are substantial gaps in **Prominence** vector scores of the control group stimulus profiles, when compared to their respective MHD profiles. Notably, all changes observed between MHD profiles and their control counterparts are statistically

significant.

When scrutinizing QS profiles from Google and Bing in tandem, we see that regardless of aggregation strategy, the range between `Positive` and `Negative`, as well as `Disorder` and `Neutral` are wider for Bing than they are for Google. From the computed stimulus scores it emerges that Bing's QS produced for MHD users are more cynical and embody more MHD indicators than Google's. When comparing M-B-QS-O, M-B-QS-BQ, and M-B-QS-R1, with C-B-QS-O, C-B-QS-BQ, and C-B-QS-R1, respectively, there are not as big of a divergence in the scores between MHD and control for the **Prominence** vector as we note for Google. Still, the differences in the **Prominence** vector scores are still at least 7 times larger between Bing's MHD and control partner of profiles. QS thus contain more terminology related MHD for MHD users, but just not as much as Google's QS did. The disparity between the scores of MHD and traditional profiles for the **Prominence** vectors persists, regardless of the SE considered, prompting us to reflect on possible causes. To an extent, the disparity makes sense, as many of the terms in the lexicon used to compute the *Intensity* vector components, overlap in the language used by both MHD and traditional users. However, the vocabulary used in **Prominence** is specifically tailored to the symptoms people with MHD experience which neurotypical individuals would not. Thus, seeing a large gap between the MHD and control profiles for **Prominence** is not unexpected.

The subliminal stimulus of QS produced by both SE have increased levels of `Anger`, `Fear`, and `Sadness` for MHD users when compared to the scores of other emotions, even more so when considering the emotion scores observed in the QS presented to traditional users. Being exposed to these bleak emotions, which are particularly prominent among top QS, could be quite damaging to users who are already in a sub-optimal head space, by triggering emotions or coping mechanism that may be unhealthy for said users.

5.1.2 The Hidden Messages of SERP

We shift our attention to the subliminal stimuli conveyed by SERP for both Google and Bing. In our analysis, we not only focus on the differences in profiles for SERP across the SE under study, but, whenever pertinent, we also make observations of discrepancies in stimulus patterns observed in QS from Section 5.1.1 and SERP.

Table 5.3: Subliminal stimuli profiles of SERP generated by Google, blue indicates significant differences between MHD profiles and their corresponding control counterparts ($p < 0.01$); purple ($p < 0.05$).

Profile	Intensity											Prominence				Evidence	
	Sentiment					Emotion						Depression		Anxiety		Disorder	Neutral
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum		
M-G-SERP-O	4.60	4.36	91.04	1.11	2.40	0.78	2.07	3.42	1.64	0.81	4.22	0.97	6.15	0.19	1.38	17.25	83.7
C-G-SERP-O	4.01	3.14	92.85	0.51	1.96	0.40	0.87	2.86	0.56	0.53	3.28	0.39	3.67	0.02	0.30	6.74	94.35
M-G-SERP-BQ	4.61	4.44	90.96	1.17	2.41	0.79	2.15	3.39	1.70	0.81	4.18	1.01	6.21	0.21	1.41	17.59	83.36
C-G-SERP-BQ	4.01	3.13	92.86	0.51	1.95	0.39	0.87	2.86	0.56	0.53	3.28	0.39	3.65	0.02	0.30	6.73	94.37
M-G-SERP-R1	4.53	4.19	91.27	1.19	2.55	0.86	2.08	3.45	1.61	0.78	4.21	0.91	6.24	0.17	1.31	15.64	85.28
C-G-SERP-R1	4.22	3.14	92.64	0.51	1.97	0.40	0.87	3.08	0.52	0.52	3.46	0.39	3.66	0.02	0.24	7.13	93.94

Table 5.4: Subliminal stimuli profiles of SERP generated by Bing, blue indicates significant differences between MHD profiles and their corresponding control counterparts ($p < 0.01$).

Profile	Intensity											Prominence				Evidence	
	Sentiment					Emotion						Depression		Anxiety		Disorder	Neutral
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum		
M-B-SERP-O	5.29	4.62	90.09	1.37	2.63	0.93	2.41	3.68	1.93	0.83	4.95	1.04	6.40	0.23	1.51	20.02	80.87
C-B-SERP-O	4.36	3.20	92.35	0.55	2.05	0.42	0.93	3.05	0.61	0.56	3.50	0.42	3.84	0.03	0.33	7.95	93.14
M-B-SERP-BQ	5.30	4.63	90.06	1.37	2.62	0.93	2.43	3.65	1.95	0.84	4.96	1.03	6.44	0.23	1.52	19.85	81.05
C-B-SERP-BQ	4.39	3.23	92.30	0.55	2.07	0.43	0.93	3.04	0.62	0.56	3.53	0.42	3.87	0.03	0.33	8.01	93.08
M-B-SERP-R1	5.12	4.36	90.52	1.81	3.09	1.29	2.99	3.50	2.45	0.86	4.45	1.06	7.76	0.24	1.56	17.26	83.65
C-B-SERP-R1	4.51	3.24	92.13	0.54	2.20	0.43	0.92	3.29	0.60	0.56	3.62	0.45	3.87	0.02	0.31	8.24	92.84

Focusing on Google's SERP (Table 5.3), we observe that M-G-SERP-R1 has significant decrease in `Disorder`, as well as increases `Neutral` with respect to the scores for the same stimuli in M-G-SERP-O and M-G-SERP-BQ. While not statistically significant change in the first ranked resource, both `Anticipation` and `Disgust` have noticeable increases. The high scores for `Disgust` and `Anticipation` on snippets positioned at the top of SERP (we treat title and snippets on SERP as snippets and refer to them as such), could be a concern: as the first thing MHD users see on a SERP, `Disgust` and `Anticipation` could be a

catalyst for unexpected behavior. This is the reason why we examine top snippets in SERP more in-depth. From the profiles of first ranked snippets, which have high scores in both `Disgust` and `Anticipation`, we observe that the majority discuss topics about sickness, sufferings, illness, sigmas, and even suicide. While suicide has a stigma surrounding it, perpetuating that stigma to users with MHD could be damaging as these users are known to struggle with suicidal ideations, by making them feel alone. Switching our attention to the MHD profiles for Bing's SERP in Table 5.4, we see an increase in bleak stimuli for top SERP snippets. In addition to `Disgust` and `Anticipation`, we find that `Anger`, `Fear`, `Sadness`, and `Despression-Forum` have significant increases over M-B-SERP-O and M-B-SERP-BQ. We further dive into what could be causing these bleak stimuli by probing sample top snippets in SERP in M-B-SERP-R1. We find the snippets tend to address topics like anxiousness, decision making, and nervousness. All these topics align with anxiety, which means that in essence Bing perpetuates anxiety by presenting MHD users who are already sensitive to such feelings snippets at the top of the page that convey a plethora of agitating emotions.

In general, the subliminal profiles of Bing SERP responding to MHD users have proportionally higher scores than Google's SERP, which lead us to believe that Bing's SERP are more stimulating than Google's for users with MHD. Other emerging trends bring to light that SERP have slightly higher `Positive` scores than `Negative` scores which we did not expect with the higher bleak emotions we saw in the `Emotion` vector. Interestingly, the inverse is true for `Positive` and `Negative` score for the MHD profiles in Section 5.1.1. Consequently, we can no longer attribute the elevated `Joy` and `Trust`, which are still present in the MHD profiles for both SE, to the negation of happy terms. Coupled with the fact that `Sadness` is no longer among the stimuli with the highest scores in the *Emotion* vector for MHD profiles, SERP are less unbalanced when it comes to contradicting

stimuli, in fact SERP may be more upbeat than QS. Furthermore, the difference in score distributions between the MHD and control profiles is substantial but is less so than it was in Section 5.1.1. For example, in Table 5.4 we see that there is no statistically significant decrease between M-B-SERP-R1 and C-B-SERP-R1 for `Joy`, i.e., `Joy` in the first ranked result on a SERP for MHD users is more aligned to that displayed for traditional users, which was not evident in Section 5.1.1. The subliminal stimuli for MHD and traditional searchers are more similar for SERP than they are for QS.

Based on the gap between the stimulus in the SERP profiles of MHD and traditional users for both SE closing, SERP seem to be less stimulating than QS, even if `Anger`, `Fear`, and `Sadness` are the stimulus that differ the most when comparing the SERP profiles of MHD and traditional users. More upbeat emotions do seem to be presented to MHD users via SERP, but the fact remains that MHD users are still exposed to less than pleasant emotions just not as predominately as they would when interacting with QS.

5.1.3 The Essence of Retrieved Resources

Nearing the end of the information seeking journey, searchers read through the content of clicked results. Retrieved resources are bigger text samples than QS or snippets, and as a result we would expect them to contain a broader range of stimulus. We begin our exploration of resources retrieve by Google by inspecting M-G-RR-O, M-G-RR-BQ, and M-G-RR-R1, along with the corresponding control profiles, C-G-RR-O, C-G-RR-BQ, and C-G-RR-R1 (in Table 5.5). In the MHD profiles of RR, we observe that the stimuli `Joy` and `Trust` have the highest scores when compared to the other stimuli in the *Emotion* vector. For the `Sentiment` vector, `Positive` is larger than `Negative`, which aligns with our findings on the `Emotion` vector. When considering M-G-RR-R1, `Anticipation`, `Joy`, and `Anxiety-Forum` have significant decreases between M-G-RR-R1 and M-G-RR-O, as well

as an increase in `Negative`. Further, `Anticipation` and `Anxiety-Forum` hav significant decreases between M-G-RR-R1 and M-G-RR-BQ.

Additionally, when contrasting the MHD user and control profiles, all observed differences in stimulus scores are significant. In the juxtaposition of Bing's profiles of RR for MHD and traditional searchers in Table 5.6, i.e., M-B-RR-O, M-B-RR-BQ, and M-B-RR-R1 versus C-B-RR-O, C-B-RR-BQ, and C-B-RR-R1, respectively), we note that, excluding the difference in `Joy` between M-B-RR-R1 and C-B-RR-R1 which are not significant, all other divergence across counterpart profiles are significant. When comparing the first ranked RR to the other aggregation strategies, we see significant decreases in `Joy`, `Trust`, and `Disorder` as well as an increase in `Neutral`. In the MHD profiles, `Anticipation`, `Joy`, and `Trust` have elevated levels over the other stimuli in the `Emotion` vector.

While Google and Bing have overlapping trend in the stimuli of RR, Bing looks to have higher stimulus levels and a spike in `Anticipation` that we do not observe among Google's RR. Neither Google's nor Bing's first ranked RR have consistently higher score in stimulus than those reported among top-10 RR, a departure from what we saw in Sections 5.1.2. In spite of the few variations of stimuli we reported in this section, for the most part RR follow the same trends as those seen for QS and SERP in Sections 5.1.1 and 5.1.2, respectively.

Table 5.5: Subliminal stimuli profiles of RR generated by Google, blue indicates significant differences between MHD profiles and their corresponding control counterparts ($p < 0.01$); purple ($p < 0.05$).

| Profile | Intensity | | | | | | | | | | | | Prominence | | | | Evidence | |
| | Sentiment | | | | Emotion | | | | | | | | Depression | | Anxiety | | Disorder | Neutral |
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust		Survey	Forum	Survey	Forum		
M-G-RR-O	4.21	3.79	91.98	0.75	1.81	0.51	1.32	2.68	1.05	0.55	3.70		0.83	5.20	0.11	1.16	8.86	92.18
C-G-RR-O	3.63	3.00	93.36	0.47	1.70	0.33	0.83	2.42	0.59	0.58	3.26		0.44	4.46	0.02	0.53	6.88	94.13
M-G-RR-BQ	4.20	3.79	91.99	0.77	1.82	0.51	1.34	2.67	1.07	0.54	3.70		0.86	5.25	0.12	1.16	9.06	91.97
C-G-RR-BQ	3.61	2.99	93.40	0.47	1.69	0.32	0.83	2.40	0.58	0.58	3.26		0.45	4.44	0.02	0.53	6.86	94.15
M-G-RR-R1	4.20	3.92	91.89	0.78	1.69	0.54	1.33	2.55	1.02	0.57	3.59		0.76	4.70	0.09	1.00	8.41	92.69
C-G-RR-R1	3.6	3.00	93.40	0.49	1.77	0.34	0.84	2.54	0.60	0.50	3.26		0.46	4.44	0.02	0.58	6.26	94.78

Table 5.6: Subliminal stimuli profiles of RR generated by Bing, blue indicates significant differences between MHD profiles and their corresponding control counterparts ($p < 0.01$).

Profile	Intensity											Prominence				Evidence	
	Sentiment			Emotion								Depression		Anxiety		Disorder	Neutral
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum		
M-B-RR-O	4.29	3.79	91.29	0.91	2.06	0.60	1.60	2.92	1.24	0.62	4.11	0.97	5.87	0.17	1.29	9.85	91.13
C-B-RR-O	3.61	3.01	93.36	0.50	1.75	0.33	0.87	2.53	0.62	0.53	3.29	0.48	4.57	0.03	0.61	6.01	95.05
M-B-RR-BQ	4.30	3.89	91.90	0.92	2.06	0.60	1.60	2.93	1.24	0.63	4.10	0.97	5.86	0.16	1.29	9.85	91.13
C-B-RR-BQ	3.61	3.02	93.35	0.50	1.76	0.33	0.87	2.52	0.62	0.54	3.31	0.48	4.60	0.03	0.61	5.97	95.09
M-B-RR-R1	4.22	3.83	91.95	0.96	1.98	0.60	1.64	2.68	1.25	0.63	3.94	0.91	5.72	0.15	1.23	8.58	92.41
C-B-RR-R1	3.64	3.00	93.33	0.50	1.80	0.33	0.89	2.63	0.64	0.52	3.38	0.48	4.53	0.03	0.56	5.83	95.25

5.2 RQ2: How do subliminal stimulus of SE responses indirectly change through the ISP for MHD users?

The starting point of the ISP is an information need expressed in the form of a search query, from which it is possible to infer the affect and any MHD indicators that users disclose to a SE through the language they use for query formulation [71, 63]. SE are an outside source of stimuli, which MHD users are known to be susceptible to [75]. Therefore, if SE were to produce responses that diverge from the original stimulus derived from users' queries it would be possible for SE to indirectly impact users' decision making process and emotional state of being [26, 56, 78] and in turn inadvertently alter these users' information seeking journeys. To fully understand the subliminal stimuli of SE responses and their potential repercussion on searchers, it is imperative to study the degree to which stimuli from SE responses fluctuate from the affective tone and the use of MHD language throughout the ISP given the opportunities that SE have to influence users through the ISP.

Consider the query "clowns" which a user may use to find an entertainer for a child's birthday party. The QS that Google produces in response to this query range from "clowns scary" to "killer klowns from outer space" to finally "clowns for hire". The query itself could be perceived as "happy" whereas the QS has an overall feeling of "fear". The first Google SERP for this query has the Wikipedia entry for clown as its top-result. The very

next result, however, is "10 famous clowns: from comical to creepy". The third result on the SERP is from the news site CNN with the title "What's with all the clowns everywhere?". The title and snippet read as inquisitive, but upon inspection of the article itself, the a user is greeted with a picture of a clown holding a machete and words like "panic", "threats", and "creepiness". What started as an interesting situation quickly turned to horrifying in a few clicks of a mouse. This example showcases that the affect of original queries do not always match that indirectly portrayed from SE responses through the ISP.

To investigate the divergence in stimuli across the ISP stages, we start by inspecting the profiles generated for user queries, as described in Section 4.5. To refer to these profiles, we use a similar naming convention as the one introduced in Section 4.3 for profiles of SE responses, but with only three letters: **{M, C}-{Q}-{O}**. We then contrast the profiles of adjacent ISP stages using the query profiles as well as some of the profiles introduced in Section 5.1. Specifically, we compare (i) the stimulus profiles of users' queries with respect to the corresponding profiles originated from QS, (ii) the stimulus profiles of users' queries with respect to the profiles elicited from the equivalent SERP, and (iii) the stimulus profiles of the collected SERP with respect to the profiles generated from the respective RR. These comparisons allow use to observe fluctuations that occur from the initiation of a search session to the presentation of RR. Much like we did in our empirical analysis for our first research question (in Section 5.1), we also consider the profiles of queries and SE responses related to traditional searchers (control group) to spur the discovery of any trends visibly *only* along the information seeking journey of MHD searchers. A high level depiction of fluctuation trends observed across the ISP for MHD and traditional searchers using both Google and Bing is captured in Figure 5.3.

Note that in the analysis reported in Section 5.1 we did not see much variation among the stimuli of profiles using different aggregations strategies. With that in mind, to stream-
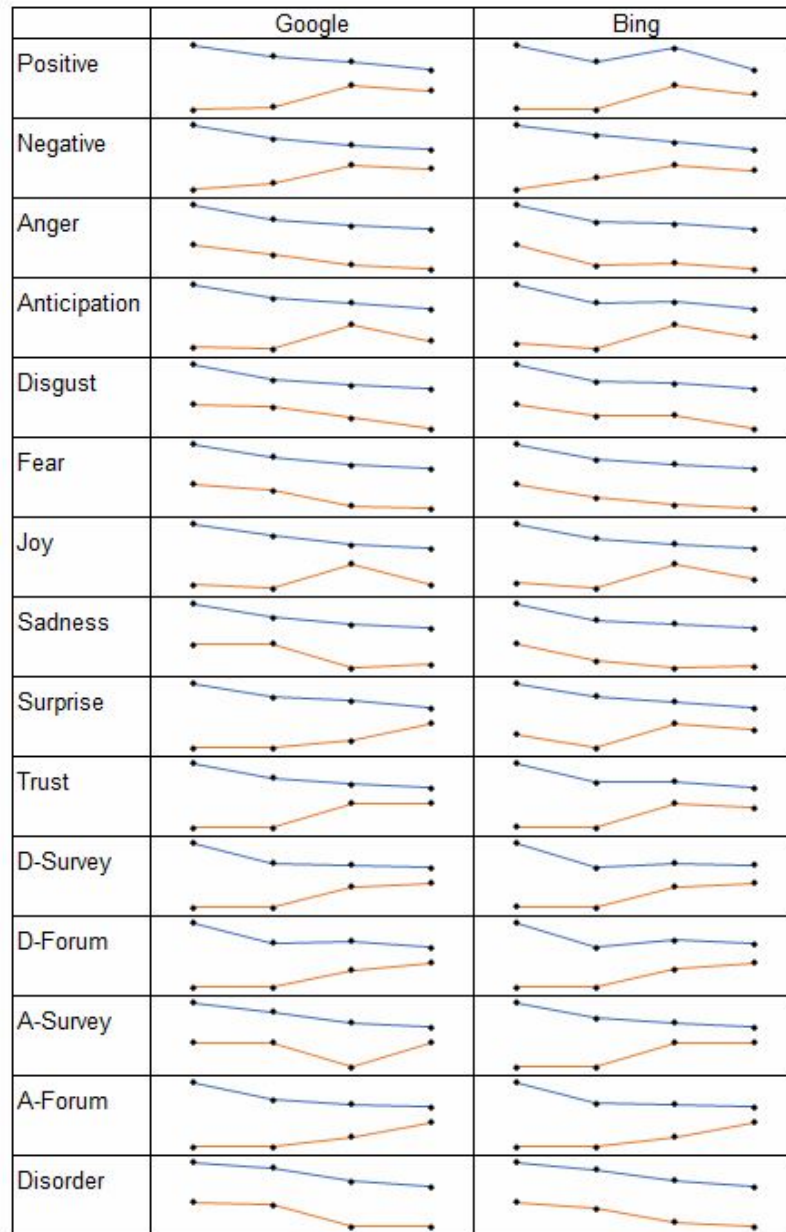
Figure 5.3: Subliminal stimulus trends across the ISP for Google and Bing. Trends lines in blue refer to MHD searchers, orange is control group. Points along each trend line represent Q, QS, SERP, and RR from left to right.

line the analysis presented in this section, we only center on the profiles generated using *overall* as the aggregation strategy. Whenever merited, we do point out notable devel-

opments resulting from profiles generated using either *by-query* or *rank-1* aggregation strategies.

5.2.1 From Queries to Query Suggestions

To establish the stimulus manifested from users' queries, we turn to the query profiles in Table 5.7. In the case of MHD searchers, from M-Q-O we see that their queries disclose higher score for `Negative` than `Positive` sentiments; there are also spikes in `Anticipation`, `Fear`, `Joy`, `Sadness`, and `Trust` in the *Emotion* vector. When comparing M-Q-O and C-Q-O, we observe a statistically significant difference in stimuli between MHD and traditional users queries. When comparing the **Prominence** vectors of MHD and control searchers, we perceive at least a 20 fold increase in vector component scores between MHD and control profiles, with the exception of `Depression-Forum`, which is only 4 times larger. Additionally, the components of the **Evidence** vector of control are at least 10 folds larger than their MHD counterparts. C-Q-O has a lower `Negative` than `Positive`, which is the opposite of what we saw in M-Q-O. As variations between M-Q-O and C-Q-O are statistically significant and there are large difference in MHD indicators, it is visible that MHD users start their information seeking journey in a very different state of mind than traditional users.

Table 5.7: Subliminal stimuli profiles of Q, blue indicates significant differences between MHD profiles and their corresponding control counterpart ($p < 0.01$).

Profile	Intensity											Prominence				Evidence	
	Sentiment			Anger	Anticipation	Disgust	*Emotion* Fear	Joy	Sadness	Surprise	Trust	*Depression* Survey	Forum	*Anxiety* Survey	Forum	Disorder	Neutral
	Pos	Neg	Obj														
M-Q-O	5.45	6.65	87.91	2.96	3.97	2.12	5.97	7.21	5.17	1.45	7.26	2.24	9.06	0.52	3.32	46.83	52.58
C-Q-O	2.47	2.33	95.16	0.69	1.59	0.49	1.23	2.41	0.75	0.51	2.43	0.11	1.98	0.02	0.15	17.08	83.08

To gauge whether with their QS SE alter in any way the stimulus conveyed in users' queries, we examine the profiles generated for users' queries vs. those generated for

associated QS by Google or Bing. Starting with Google, from M-Q-O and M-G-QS-O in Table 5.8 we observe that differences in scores in the profiles for QS vs Q are statistically significant except for `Anxiety-Survey` and `Positive`. All scores in M-G-QS-O, except `Objective` and `Disorder`, decrease when compared to their counterparts in M-Q-O, but M-G-QS-O still follow a similar distribution of scores in M-Q-O. We notice that the stimulus scores in M-G-QS-R1 are closer to those of M-Q-O, meaning the top-ranked QS is the closest to the affect of a users query. This shows that while Google aligns with the stimuli of the queries of MHD users, it is muting the stimulus potency.

Table 5.8: Subliminal stimuli profile of Q along with counterpart profile for QS generated by Google, blue indicates that profile components in Q differ significantly from the respective ones in QS ($p < 0.01$).

| Profile | Intensity | | | | | | | | | | | | Prominence | | | | Evidence | |
| | Sentiment | | | Emotion | | | | | | | | | Depression | | Anxiety | | | |
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum	Disorder	Neutral
M-Q-O	5.45	6.65	87.91	2.96	3.97	2.12	5.97	7.21	5.17	1.45	7.26	2.24	9.06	0.52	3.32	46.83	52.58
M-G-QS-0	4.87	5.00	89.29	1.67	2.73	1.10	3.62	4.93	2.86	0.94	5.10	1.08	5.98	0.37	1.78	38.03	61.91
C-Q-O	2.47	2.33	95.16	0.69	1.59	0.49	1.23	2.41	0.75	0.51	2.43	0.11	1.98	0.02	0.15	17.08	83.08
C-G-QS-O	2.63	2.53	90.40	0.60	1.56	0.47	1.12	2.34	0.76	0.51	2.48	0.10	2.18	0.02	0.15	15.81	84.60

The tempering of stimuli from M-Q-O to M-G-QS-O could be the norm for Google, i.e., observed also among Q to QS interactions among traditional users. We explore distribution trends between C-Q-O and C-G-QS-O (i.e., the QS and Q profiles for the control group) seeking to confirm if they remain the same as the ones detected for MHD searchers. From Table 5.9 it is apparent that the profile scores that significantly differ are `Objective` and `Disorder` which decrease, as well as `Neutral` which increases. The profile scores in C-Q-O, except for `Positive`, *Negative*, `Trust`, `Depression-Forum`, and `Neutral` increase when compared to scores computed for C-G-QS-O. The magnitude of these increases, however, is far less those observed between M-Q-O and M-G-QS-O. For instance, between C-Q-O and C-G-QS-O there are slight but not significant changes to no changes at all in `Sadness`, `Surprise`, `Depression-Survey`, `Anxiety-Survey`, and `Anxiety-Forum`, a

phenomena not seen between M-Q-O and M-QS-O. In the end, the differences we see in the shift from Q to QS in Google are not typical as there are difference in the stimuli SE response convey to MHD versus control users, with the affect from MHD queries being so deaden in the stimulus of QS project.

The algorithm Bing uses to produce QS could differ from Google's and thus result in disparate changes from Q to QS. By comparing M-Q-O and M-B-QS-O profiles (in Table 5.9) we look for possible shifts in the stimulus that users communicate via their queries vs. the stimulus of Bing's QS. One key dissimilarity from M-Q-O to M-B-QS-O is the significant decrease in `Anxiety-Survey`; this is something not observed in Google's transitions from Q to QS. Looking at the remaining profile scores, we see many of the same trends identified when comparing M-Q-O and M-G-QS-O. M-B-QS-O has lower stimulus scores than M-Q-O, but still maintains the general pattern of stimuli scores observed in M-Q-O. Collectively, it appears that much like Google, the subliminal stimulus in Bing's QS align with affect in users queries but dampens the stimulus strength.

Table 5.9: Subliminal stimulus profile of Q along with the counterpart profile for QS generated by Bing. Blue indicates significant differences of profile components for Q with respect to QS ($p < 0.01$); purple ($p < 0.05$).

Profile	Intensity											Prominence				Evidence	
	Sentiment			Anger	Emotion							Depression		Anxiety		Disorder	Neutral
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum	Disorder	Neutral
M-Q-O	5.45	6.65	87.91	2.96	3.97	2.12	5.97	7.21	5.17	1.45	7.26	2.24	9.06	0.52	3.32	46.83	52.58
M-B-QS-0	4.69	5.44	89.83	1.50	2.54	1.03	3.26	4.73	2.58	1.03	4.84	0.78	5.36	0.3	1.59	35.27	64.58
C-Q-O	2.47	2.33	95.16	0.69	1.59	0.49	1.23	2.41	0.75	0.51	2.43	0.11	1.98	0.02	0.15	17.08	83.08
C-B-QS-O	2.37	2.73	94.84	0.53	1.47	0.41	1.03	2.22	0.65	0.45	2.37	0.1	2.02	0.02	0.15	13.97	86.40

We once again turn to our control group, via C-Q-O and C-B-QS-O, to assess if the QS produced for MHD users have different stimulus than those generated for traditional users. is indeed the case. There are significant changes seen between C-Q-O and C-B-QS-O for the **Evidence** vector The differences noted between traditional Q and Bing's QS, however, mimic those observed in Google, with the only exception being Google's QS

have a decrease in `Objective` but no increase in `Negative`. While it appears that the diminishing of stimuli is not the norm for Bing's QS, it is more common than it is for Google's QS.

Based on our comparison of Q and QS, we note that distributions of stimuli on QS profiles for both SE align with the those for users' queries but dull the stimulus potency. The divergence from users expressed affect and MHD indicators could alter how users interact with the SE. If users submit very emotionally-charged queries and all the suggestions the SE provides are less emotional, users could lean towards QS influenced by the emotion dimension of relevance and choose one that is not necessarily the one that best captures their search intent (a phenomena well-documented in emotion-aware recommender system literature [53]). We surmise that that by dulling the stimulus in QS the SE could keep MHD users from finding support (for example song lyrics they find comforting) or instead lead them stray towards SERP that contain triggers for their MHD.

5.2.2 From Queries to SERP

We next peruse the transition from queries to SERP, we compare M-Q-O to M-G-SERP-O (Table 5.10). We see a significant decrease in all stimuli except for `Objective` and `Neutral`, which increase. There is a noticeable gap between the scores of the stimulus of M-Q-O and M-G-SERP-O implying that the stimulus Google conveys to MHD users with its SERP is diverging from the stimulus of a users query. When considering Bing's SERP responses to users queries (M-B-SERP-O c.f. C-B-SERP-O in Table 5.11), emerging trends for the most part closely resemble Google's. We note that when comparing M-B-SERP-O and M-Q-O all changes in stimulus scores are statistically significant, except `Positive` and `Negative`. Both Google's and Bing's SERP when compared with MHD users queries display similar stimuli, but the ferocity with which they appear is greatly diminished from

the users original affect, which causes a notable disparity between a users mental state and the responses they are being provided.

Table 5.10: Subliminal stimulus profile of Q along with the counterpart profile for SERP generated by Google. Blue indicates significant differences of profile components for Q with respect to SERP ($p < 0.01$).

Profile	Intensity												Prominence				Evidence	
	Sentiment					*Emotion*							*Depression*		*Anxiety*		Disorder	Neutral
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust		Survey	Forum	Survey	Forum		
M-Q-O	5.45	6.65	87.91	2.96	3.97	2.12	5.97	7.21	5.17	1.45	7.26		2.24	9.06	0.52	3.32	46.83	52.58
M-G-SERP-0	4.60	4.36	91.04	1.11	2.40	0.78	2.07	3.42	1.64	0.81	4.22		0.97	6.15	0.19	1.38	17.25	83.7
C-Q-O	2.47	2.33	95.16	0.69	1.59	0.49	1.23	2.41	0.75	0.51	2.43		0.11	1.98	0.02	0.15	17.08	83.08
C-G-SERP-O	4.01	3.14	92.85	0.51	1.96	0.40	0.87	2.86	0.56	0.53	3.28		0.39	3.67	0.02	0.30	6.74	94.35

Table 5.11: Subliminal stimulus profile of Q along with the counterpart profile for SERP generated by Bing. Blue indicates significant differences of profile components for Q with respect to SERP ($p < 0.01$); purple ($p < 0.05$).

Profile	Intensity												Prominence				Evidence	
	Sentiment					*Emotion*							*Depression*		*Anxiety*		Disorder	Neutral
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust		Survey	Forum	Survey	Forum		
M-Q-O	5.45	6.65	87.91	2.96	3.97	2.12	5.97	7.21	5.17	1.45	7.26		2.24	9.06	0.52	3.32	46.83	52.58
M-B-SERP-0	5.29	4.62	90.09	1.37	2.63	0.93	2.41	3.68	1.93	0.83	4.95		1.04	6.40	0.23	1.51	20.02	80.87
C-Q-O	2.47	2.33	95.16	0.69	1.59	0.49	1.23	2.41	0.75	0.51	2.43		0.11	1.98	0.02	0.15	17.08	83.08
C-B-SERP-O	4.36	3.20	92.35	0.55	2.05	0.42	0.93	3.05	0.61	0.56	3.50		0.42	3.84	0.03	0.33	7.95	93.14

We study C-Q-O and C-G-SERP-O, as well as C-Q-O and C-B-SERP-O, to get a sense for the behavior of the SERP present to traditional versus MHD users. C-Q-O has significantly higher scores than C-G-SERP-O for `Positive`, `Negative`, `Anticipation`, `Joy`, `Trust`, `Depression-Survey`, `Depression-Forum`, `Anxiety-Forum`, and `Neutral`; it also has lower scores for `Objective`, `Anger`, `Fear`, `Sadness`, and `Disorder`. These changes deviate heavily from the changes we detect between M-Q-O to M-G-SERP-O in the amount of stimuli that had significant changes. The changes to stimuli for MHD searches are mostly decreases in stimulus scores, while for traditional searchers bleak emotions have decrease and upbeat ones have increases. When exploring Bing's response behaviour for the control group we see that with the exception of `Disgust`, `Surprise`, and `Anxiety-Survey`, all changes in stimulus are statistically significant for C-Q-O vs.

C-B-SERP-O. Additionally, we see the same elevated of upbeat and decreasing of bleak emotions in Bing that was present in Google.

From findings arising as a result of analysing transitions in stimulus from queries to SERP on MHD and control profiles, we deduce that even though both SE produce SERP with dissimilar stimulus for queries to both MHD and traditional searchers. However, how the specific fluctuations in stimuli from Q to SERP depend upon the user who initiates the search process. In the case of traditional users, SERP are more upbeat and less bleak than the originating queries, resulting in a SERP that conveys emotional stability. Instead, for MHD users, SERP stimuli are deadened in respect to those encapsulated in their queries, thus causing spikes in bleak stimuli in SERP to stand out more. To illustrate the difference in the change in stimulus between queries and SERP for MHD and traditional users, think about queries as dark humor: they are bleak but kind of funny. If a bit of dark and funny is removed from dark humor, all that is left is a mildly dark statement, which is the equivalent to the SERP generated from the original query. We infer that this could leave MHD users in a very different place emotionally than when they start their information seeking journey.

5.2.3   From SERP to RR

We have discussed in previous sections (Sections 5.2.1 and 5.2.2), how the subliminal stimulus of QS and SERP have mutated the affect and MHD indicators users express in their queries, but we have not considered how SE represent the stimuli of web content through SERP. To investigate the difference in stimuli between the snippets on SERP to the RR they represent, we compare Google's SERP and RR profiles (M-G-SERP-O and M-G-RR-O, respectively) in Table 5.12. The changes in stimulus between SERP and RR for MHD users have significant decreases, except for `Objective` and `Neutral`. When considering Bing instead, using the profiles M-B-SERP-O and M-B-RR-O in Table 5.13,

we see a very similar picture to what we saw with Google's SERP and RR. However, we do note that Bing does not have a significant change in `Depression-Survey` and has a larger gap in the scores for `Positive` between RR and SERP than Google does. The gap in stimulus between SERP and RR implies that for both SE, SERP are amplifying the stimulus seen in RR for MHD users, which can be problematic when MHD users are selecting resources on emotional subjects.

Table 5.12: Subliminal stimuli profiles of SERP and RR generated by Google, blue indicates significant differences between SERP and RR ($p < 0.01$); purple indicates ($p < 0.05$).

Profile	Intensity											Prominence				Evidence	
	Sentiment					Emotion						Depression		Anxiety			
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum	Disorder	Neutral
M-G-SERP-O	4.60	4.36	91.04	1.11	2.40	0.78	2.07	3.42	1.64	0.81	4.22	0.97	6.15	0.19	1.38	17.25	83.7
M-G-RR-O	4.21	3.79	91.98	0.75	1.81	0.51	1.32	2.68	1.05	0.55	3.70	0.83	5.20	0.11	1.16	8.86	92.18
C-G-SERP-O	4.01	3.14	92.85	0.51	1.96	0.40	0.87	2.86	0.56	0.53	3.28	0.39	3.67	0.02	0.30	6.74	94.35
C-G-RR-O	3.63	3.00	93.36	0.47	1.70	0.33	0.83	2.42	0.59	0.58	3.26	0.44	4.46	0.02	0.53	6.88	94.13

Table 5.13: Subliminal stimuli profiles of SERP and RR generated by Bing, blue indicates significant differences between SERP and RR ($p < 0.01$).

Profile	Intensity											Prominence				Evidence	
	Sentiment					Emotion						Depression		Anxiety			
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum	Disorder	Neutral
M-B-SERP-O	5.29	4.62	90.09	1.37	2.63	0.93	2.41	3.68	1.93	0.83	4.95	1.04	6.40	0.23	1.51	20.02	80.87
M-B-RR-O	4.29	3.79	91.29	0.91	2.06	0.60	1.60	2.92	1.24	0.62	4.11	0.97	5.87	0.17	1.29	9.85	91.13
C-B-SERP-O	4.36	3.20	92.35	0.55	2.05	0.42	0.93	3.05	0.61	0.56	3.50	0.42	3.84	0.03	0.33	7.95	93.14
C-B-RR-O	3.61	3.01	93.36	0.50	1.75	0.33	0.87	2.53	0.62	0.53	3.29	0.48	4.57	0.03	0.61	6.01	95.05

While SERP are not accurately representing the stimuli of RR for MHD users, it has yet to be seen if this is the case for traditional users. Considering C-G-SERP-O and C-G-RR-O, we immediately notice that there are more stimuli that do not have significant changes than MHD users profiles. there are not significant changes in `Fear`, `Sadness`, `Trust`, `Anxiety-Survey`, and `Disorder`. Additionally, the magnitude of changes between SERP and RR profiles are smaller for traditional users than for MHD. When examining C-B-SERP-O and C-B-RR-O, we see less significant changes occurring with `Sadness`, `Surprise` and `Anxiety-Survey` not significantly changing, as well as a difference in the expanse in changes between SERP and RR for MHD vs. traditional users. We also see that

the different aggregation strategies varies in the amount of significant changes, with overall having the most changes and by-query having the least. From these observations it seems there are less discrepancies between RR and their SERP representations for traditional users than there are for MHD users, thus SERP are more representative of RR in general when responding to traditional users.

SERP display a heightened level of the stimuli with respect to that conveyed in RR; more so for RR responding to MHD users than traditional ones. Consider the query "trypophobia" (trypophobia is the fear of clusters of small holes). For a user looking for information about the phobia, a SERP snippet where the stimuli does not match that of the corresponding RR may not be a concern. Contrarily, individuals experiencing said phobia and turning to a SE to look for support, may be drawn to snippets with high levels of fear, in the hopes of finding validation of their experiences. Unfortunately, there exists pages displaying content mocking the phobia and therefore could contain triggering terminology; if these resources are misrepresented on their snippet, they may draw the attention of users with the phobia and trigger them. In this particular case, SE inadvertently emulate the actions of click bait rather than acting as an unbiased provider of relevant resources. Alternatively, a SERP snippet that has high stimuli for the corresponding RR could deter MHD users from clicking on that resource, due to the fear of encountering triggers, even though the RR is relevant to their information need.

5.3   Discussion

Outcomes from the empirical analysis of our research questions reveal several interesting phenomena that occur when MHD searchers interact with SE. (Trend highlights of RQ1 and RQ2 are illustrated in Figures 5.1 and 5.3, respectively) We first examined each of

the ISP stages as snapshots (RQ1), allowing us to see that each SE responds to users with MHD with elevated levels of `Anger`, `Fear`, and `Sadness`. These findings are concerning, as prolonged exposure to such cynical emotions can be damaging to the mental health of users with MHD [33]. Additionally, traditional searches are not exposed to elevated levels of bleak emotions but tend to have more neutral stimulus levels, showcasing that SE responses are different depending on user type. How inflated cynical emotions become for MHD users depends on which ISP stage the user is at; The inflating of emotions is most prominent in QS and the least in RR. What rank a SE response is impacts the potency of stimuli that is conveyed to users. Top-ranked responses consistently display slightly more intense emotions when contrasted with scores computed as a result of averaging all offered choices for QS, SERP, or RR. Thus, MHD users are faced with cynical emotions at the very first responses they read. At the very least a user usually needs to read the first snippet on a SERP to know if their query produced relevant resources. While this would not be troubling for traditional users, given the stimulus present in their SERP, being forced to read such snippets could be harmful to users with MHD. We investigated further into top-ranked responses, as we noticed that they could boost detrimental combinations of emotions, recall our dive into `Anticipation` and `Disgust` in Section 5.1. Finding imply that top-ranked responses tend to lead to or contain language related to topics that can be triggering for users with MHD. Exposing MHD users to triggers is not only detrimental to their health, but it can also be dangerous. Both SE under study display the issues we have set forth. Nevertheless, Bing usually has higher stimulus scores than Google's. We see patterns such as the ones we have discussed in responses to MHD users happening at every stage of the ISP stage, which could wear on MHD users mentally, causing unneeded strain on their already compromised mental health.

In answering RQ1 we noticed that emerging trends in stimulus responses at ISP stages

appeared in all stages of the ISP. Significant distinctions we have found when comparing MHD and control profiles of both Google's and Bing's ISP stages confirm that a steady decrease in the differences in stimulus scores between MHD and control profiles is transpiring. The same can be said about fluctuations in stimuli displayed by first ranked text samples. Additionally, most of the high scores in bleak emotions we have seen in QS and SERP subsided, and instead upbeat emotions increase for RR. However, if we consider the difference in specific stimuli between MHD and control profiles, `Anger`, `Disgust`, `Fear`, and `Sadness` have the largest differences, manifesting that traditional users are not subjected to these elevated levels of cynical emotions.

Impelled by the fact that the ISP stages are not done in isolation but are in fact a process, we turned our attention to how stimuli explicitly differ from one ISP stage to the next (RQ2). When providing information to MHD users, the emotions and MHD indicators they express at the time of initiating the search can influence the way they internalize information, making it vital that we understand the stimulus RR present to users. RR have the most stable profiles among those generated for different types of SE responses and the least differences between subliminal stimulus profiles of traditional and MHD users when compared to QS and SERP. However, users with MHD are still presented with proportionally higher cynical emotions than their traditional counterparts in RR.

Through our analysis, we have seen how the affect in a user's original query gets distorted gradually through the stages of the ISP. While for traditional users it seems as though the stimuli presented in queries is stabilizing in SE responses, for MHD users their expressed affect and MHD indicators id dampened. The best way to illustrate what is happening in the ISP stage is to visualize a funhouse mirror. While these mirrors reflex back what is in front of them, it returns a distorted vision of what it saw. For traditional users they get back a more stable version of what they started with. Unfortunately, MHD

users get back an image with less contrast, the dark is are not so dark, but the brights are also not as bright. The funhouse mirror effect can be especially troubling when MHD users seek some form of validation through SE, as rather than finding resources presented in a similar mindset, they find placating versions of their emotions.

Leveraging the insights gained from our analysis associated with RQ1 and RQ2, we capture a current snapshot of how Google and Bing handle inquiries initiated by MHD searchers, progressing forward in answering the question: *How do SE respond to users with MHD?*. Let us consider a search session initiated by a user named *Kristoff*, who has a major depressive disorder. From RQ2 we know that Kristoff's query is likely to contain high levels of fear and sadness, as well as be prominent in depressed terminology and other MHD indicators. The first SE response Kristoff is exposed to is QS, which we know from RQ2 will have similar but diluted emotions to his query as well as less MHD terminology but it will still be rather present. We also know from RQ1 that the QS will have elevated levels of anticipation on top of the emotions already present from Kristoff's query and the first QS he is meet with will be the most stimulating. It is possible that, based on our findings, Kristoff is subjected to new emotions and varying terms associated with MHD than when he started the search session. The SE propagating MHD-indicative terms introduces a new issue. One of the most stigmatized topics related to depression is suicide and many individuals use SE for information on suicide [70]. Studies have demonstrated that the portrayal of suicide in media and online content can increase suicide rates [17] as well as how suicide-related information retrieved from SE can have either a positive or negative impact on searchers [70]. Given that Kristoff has major depressive disorder, a disorder for which thoughts of suicide is a symptom [5], our findings coupled with those from the aforementioned studies suggests that by projecting MHD indicators to MHD users, SE can exacerbate some symptoms of MHD, in the case of Kristoff suicidal

tendencies. After Kristoff has finalized his query and moved to SERP, we know from RQ1 and RQ2 that he will be meet again with a snippet at the top of the SERP that has the highest levels of emotions carried over from QS, compared to the rest of the page. Along with dulling of emotions from what he experienced in QS, the changing in the portion of MHD terms changes. Once Kristoff, who likely had his decision-making process altered due to the shifting of emotions [75], selects a RR based on a snippet, he will move to a RR. From RQ2 we know that the RR will have diluted emotions from what the snippet presented on the SERP, which if Kristoff had picked the RR expecting it to have the same level of emotions as displayed as on the SERP he may be disappointed. At the end of Kristoff's search session, he may or may not have found what he was looking for but, he has been exposed to a variety of affects and MHD indicators, which could exacerbate the symptoms of his MHD or change his current state of mind.

We use a sample MHD user's search session, which touches on each stage of the ISP, to take our empirical findings from theory to practice. With this, we have not only conveyed how SE respond to users with MHD but have also shown the possible consequences of a user with a MHD interacting with a persuasive technology like SE. In the process, we have uncovered important knowledge gaps that researchers in the Information Retrieval community should leverage in their quest to improve search systems that better serve all user groups, not just traditional ones.

CHAPTER 6

CONCLUSION, LIMITATIONS, AND FUTURE WORK

In this manuscript, we discuss the empirical analysis we conducted in order to understand how Google and Bing respond to users with anxiety and depression. We use affect and MHD indicators as lenses driving our exploration. Specifically, we consider the subliminal stimulus that each of the search engines (SE) under study portray at different stages of the information seeking process (ISP) as well as how stimuli fluctuate from one stage to another. Along the way, we consider the subliminal stimulus SE present to traditional users, i.e., "average" searchers for whom SE are designed. These users serve as a control group, enabling us to put into perspective if the subliminal stimulus of SE responses are the result of the SE algorithmic design or if they correlate with the user group that initiates the search session.

Even though some of the outcomes emerging from our analysis were anticipated due to the affect and language associated with the MHD under study, the vast majority of the findings arising from our empirical explorations reveal new insights into the subliminal stimulus of SE. Most notably, the stimulus presented to users with MHD has bleak affects and MHD indicators are elevated when initiating a search then decline throughout the ISP. The same is not true of the experiences of traditional users' interactions with SE, as the stimuli conveyed to them is more balanced emotionally and contains far fewer indicators of MHD. Additionally, upbeat emotions are elevated in SE responses to MHD users, but

due to the high score in negative sentiment, this is probably a misrepresentation of the actual emoticons expressed. Our findings also indicate that first-ranked text samples (be that query suggestions, snippets, and retrieved resources) are the most stimulating at every stage of the ISP when compared to the text of top-ten responses. Therefore, the very first response users read is the most stimulating and thus the most likely to alter users' state of mind.

Results of our work align with some prior research which has highlighted the sentiment and emotions of SE results when responding to traditional users. The subliminal stimulus profiles of Bing's SERP convey to traditional users echo those described in [37]. Moreover, the study introduced in this thesis extends that of Kazai et al. [37], who create emotional profiles of search engines for traditional searchers, by including MHD indicators and emotions, as well as considering additional SE functionality and other commercial SE. Lessons learned from our exploration are also in line with those demonstrated by Landoni et al. [39], who examined the affect of SE responses to children, regarding the need for considering the SE design and evaluating the affective dimension. To the best of our knowledge, the analysis we presented in this thesis is novel as even though literature exists that examine the affect response of SE responses, it is solely focused on traditional users. Instead, we advance knowledge by considering MHD users, MHD indicators, and all stages of the ISP.

Using the knowledge gained for this work, the Information Retrieval (IR) community could expand its look into demographic factors in search systems– past age and gender which are already prevalent in the literature [47]. We envision that the preliminary outcomes resulting from our study can bring to attention the need to also consider searchers with MHD and their needs in the design, development, and evaluation of search tools. Given how in-vogue fairness research in machine-learning and artificial-intelligence is [15, 30], and justifiably so, it is excepted that more research would be done on the al-

gorithmic side of SE for MHD searchers. Particularly given that research efforts thus far have primarily showcased biases of gender [13], or even clinical terms [81], in text but little attention that has been paid to MHD or text specific to SE which is worrying in light of the outcomes and implications enumerated in our thesis. Based on the results of our analysis, it is imperative that the terminology used by MHD searchers becomes part of search algorithms so as to address terminology and other language triggers of MHD users. The emotional component is one of the many that can spur search inquires [73], and from our results, it seems that MHD can also be a driving force in motivating inquires given the prominence of MHD indicators in our MHD user queries. Additionally, studies into the role that SE response stimuli play in influencing users' behavior, have been limited [4]. The analysis we have presented can act as a first step in propelling forward research on SE and user behavior on an emotional level. From a theoretical perspective, what we have observed in terms of emotions in the ISP aligns with the idea set forth by Kuhlthau [38], that indeed affect is important throughout the stages of the ISP. Leveraging lessons learned from our study, further studies could expand theoretical frameworks for the ISP by integrating more stimuli cues from both a system and user perspective. Additionally, our work has implications on the necessity to update how the IR community defines information need and search intent. Both these core concepts could be extended to include an affective dimension, or even an MHD indicator dimension, as emotions impact the decision-making of users thus altering their needs and intentions in search sessions and it is clear from our analysis that SE convey emotions to users consistently.

As with any research study, we identify some limitations. Although we consider the two most popular commercial search engines, they are not the only ones that users with MHD turn to. It would be interesting to expand this work by conducting a similar analysis based on different languages and locations, as MHD are expressed and treated differently in other

cultures. Expanding the users analyzed could also take the form of users with other forms of MHD. In this analysis, we only study anxiety and depression to not overgeneralize, but users with other MHD, like post-traumatic stress disorder, could also be exposed to harmful stimuli through SE responses but there is no guarantee that responses would be the same, making other MHD important to examine. We also know that how MHD present in children is different than adults, so considering children is another path for future exploration. One of the hurdles to expanding the current exploration is data. We created synthetic datasets to enable our analysis, however, it would be worth dedicating resources to collecting real query logs from users, as this would enable us to get a more fine-grained look at how SE respond to users. Additionally, while we built lexicons for our analysis, lexicons can not encompass every aspect of MHD, which is why we also considered state-of-the-art natural language processing strategies as well. However, since existing strategies are designed and trained with text from social media platforms, which differs for the type of text seen in queries, snippets, and web resources, there is no guarantee they will generalize to the point of detecting MHD in SE. Thus, future explorations of MHD detection strategies for SE that do not require query logs is another possible avenue for future research.

Upon reflecting on the outcomes of our empirical analysis, we unearthed open research directions. Is the way SE responded good for users with MHD? How should SE respond to these users? While most of our profiles had significant changes between MHD and traditional users, do users notice that the changes in the stimuli being resented to them? SE have become the first place most people look for information about almost anything. If a user with a MHD starts searching for information related to their symptoms, what if they are meet with triggering information or given access to the ideas that will ultimately lead to them harming themselves? The use of SE has in a way removed the stopgap between people and harmful information, as people used to have to talk to someone to get it, who

in turn could act and redirect a person in an attempt to help them. The only stopgap that exists for SE is the suicide prevention hotline that appears when a quire directly references the harming of one's self. While providing the hotline shows an effort it does not account for any other indicators that a user is suffering.

SE have neither the capacity to alleviate the anxiety or depression (or any other MHD for that matter) users are afflicted by nor prevent searchers from ending their lives just by the inherent nature of the results or suggestions SE present if that is what these searchers truly wish. Nevertheless, it is our duty as researchers to do what we can to support the users our algorithms serve. We aspire for this work to set the foundation for more research into SE and MHD users.

REFERENCES

[1] 99firms.com. Search engine statistics. Available at: `https://99firms.com/blog/search-engine-statistics/`, 2019. (accessed March 20, 2020).

[2] Mickeel Allen. Technological influence on society. Available at: `https://www.bctv.org/2019/11/07/technological-influence-on-society/`, 2019. (accessed August 21, 2020).

[3] Mental Health America. The state of mental health in america. Available at: `https://www.mhanational.org/issues/state-mental-health-america`, 2020. (accessed August 18, 2020).

[4] Ioannis Arapakis, Joemon M Jose, and Philip D Gray. Affective feedback: an investigation into the role of emotions in the information seeking process. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 395–402, 2008.

[5] American Psychiatric Association et al. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[6] American Psychological Association. Apa dictionary of psychology. Available at: `https://dictionary.apa.org/affect`, 2020. (accessed August 21, 2020).

[7] Agata MP Atayde, Sacha C Hauc, Lily G Bessette, Heidi Danckers, and Richard Saitz. Changing the narrative: a call to end stigmatizing terminology related to substance use disorders, 2021.

[8] John W Ayers, Eric C Leas, Derek C Johnson, Adam Poliak, Benjamin M Althouse, Mark Dredze, and Alicia L Nobles. Internet searches for acute anxiety during the early stages of the covid-19 pandemic. *JAMA Internal Medicine*, 180(12):1706–1707, 2020.

[9] Ion Madrazo Azpiazu, Nevena Dragovic, Oghenemaro Anuyah, and Maria Soledad Pera. Looking for the movie seven or sven from the movie frozen?: A multi-perspective strategy for recommending queries for children. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR)*, pages 92–101, 2018.

[10] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, volume 10, pages 2200–2204, 2010.

[11] Amanda Baker, Naomi Simon, Aparna Keshaviah, Amy Farabaugh, Thilo Deckers-bach, John J Worthington, Elizabeth Hoge, Maurizio Fava, and Mark P Pollack. Anxiety symptoms questionnaire (asq): development and validation. *General Psychiatry*, 32(6), 2019.

[12] Aaron T Beck, Calvin H Ward, Mock Mendelson, Jeremiah Mock, and John Erbaugh. An inventory for measuring depression. *Archives of general psychiatry*, 4(6):561–571, 1961.

[13] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too

big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623, 2021.

[14] Jared D Bernard, Jenna L Baddeley, Benjamin F Rodriguez, and Philip A Burke. Depression, language, and affect: an examination of the influence of baseline depression and affect induction on language. *Journal of Language and Social Psychology*, 35(3): 317–326, 2016.

[15] Sarah Bird, Krishnaram Kenthapadi, Emre Kiciman, and Margaret Mitchell. Fairness-aware machine learning: Practical challenges and lessons learned. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 834–835, 2019.

[16] Michael Birnbaum, Sindhu Kiranmai Ernala, Asra Rizvi, Munmun De Choudhury, and John Kane. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. *Journal of medical Internet research*, 19(8):e289, 2017.

[17] Jeffrey A Bridge, Joel B Greenhouse, Donna Ruch, Jack Stevens, John Ackerman, Arielle H Sheftall, Lisa M Horowitz, Kelly J Kelleher, and John V Campo. Association between the release of netflix's 13 reasons why and suicide rates in the united states: An interrupted time series analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, 59(2):236–243, 2020.

[18] Andrew Campbell, Brad Ridout, Melina Linden, Brian Collyer, and John Dalgleish. A preliminary understanding of search words used by children, teenagers and young adults in seeking information about depression and anxiety online. *Journal of Technology in Human Services*, 36(4):208–221, 2018.

[19] Sergiu Chelaru, Ismail Sengor Altingovde, Stefan Siersdorfer, and Wolfgang Nejdl. Analyzing, detecting, and exploiting sentiment in web queries. *ACM Transactions on the Web (TWEB)*, 8(1):1–28, 2013.

[20] Paula Glenda Ferrer Cheng, Roann Munoz Ramos, Jó Ágila Bitsch, Stephan Michael Jonas, Tim Ix, Portia Lynn Quetulio See, and Klaus Wehrle. Psychologist in a pocket: lexicon development and content validation of a mobile-based app for depression screening. *JMIR mHealth and uHealth*, 4(3):e88, 2016.

[21] J. Clement. U.s. internet usage penetration 2019, by age group. Available at: `https://www.statista.com/statistics/266587/percentage-of-internet-users-by-age-groups-in-the-us/`, 2019. (accessed March 20, 2020).

[22] Mark É Czeisler, Rashon I Lane, Emiko Petrosky, Joshua F Wiley, Aleta Christensen, Rashid Njai, Matthew D Weaver, Rebecca Robbins, Elise R Facer-Childs, Laura K Barger, et al. Mental health, substance use, and suicidal ideation during the covid-19 pandemic—united states, june 24–30, 2020. *Morbidity and Mortality Weekly Report*, 69(32):1049, 2020.

[23] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, 2013.

[24] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, page 2098–2110, 2016. doi: 10.1145/2858036.2858207.

[25] Gianluca Demartini and Stefan Siersdorfer. Dear search engine: What's your opinion about...? sentiment analysis for semantic enrichment of web search results. In *Proceedings of the 3rd International Semantic Search Workshop*, 2010. URL `https://doi.org/10.1145/1863879.1863883`.

[26] Jan B Engelmann, Todd A Hare, Andrew S Fox, Regina C Lapate, Alexander J Shackman, and Richard J Davidson. Emotions can bias decision-making processes by promoting specific behavioral tendencies. 2018.

[27] Reuben Fischer-Baum. What 'tech world' did you grow up in? Available at: `https://www.washingtonpost.com/graphics/2017/entertainment/tech-generations/`, 2017. (accessed August 21, 2020).

[28] BJ Fogg. *Persuasive technology: using computers to change what we think and do*. Morgan Kaufmann, 2003.

[29] Tatiana Gossen. *Search engines for children: search user interfaces and information-seeking behaviour*. Springer, 2016.

[30] Anhong Guo, Ece Kamar, Jennifer Wortman Vaughan, Hanna Wallach, and Meredith Ringel Morris. Toward fairness in ai for people with disabilities: A research roadmap. *arXiv preprint arXiv:1907.02227*, 2019.

[31] Max Hamilton. The assessment of anxiety states by rating. *British Journal of Medical Psychology*, 32(1):50–55, 1959.

[32] Max Hamilton. Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, 6(4):278–296, 1967.

[33] Markham Heid. You asked: Is it bad for you to read the news constantly? Available at: `https://time.com/5125894/is-reading-news-bad-for-you/`, 2018. (accessed March 18, 2020).

[34] Michael Herdman, Claire Gudex, Andrew Lloyd, MF Janssen, Paul Kind, David Parkin, Gouke Bonsel, and Xavier Badia. Development and preliminary testing of the new five-level version of eq-5d (eq-5d-5l). *Quality of Life Research*, 20(10): 1727–1736, 2011.

[35] Nicholas C Jacobson, Damien Lekkas, George Price, Michael V Heinz, Minkeun Song, A James O'Malley, and Paul J Barr. Flattening the mental health curve: Covid-19 stay-at-home orders are associated with alterations in mental health search behavior in the united states. *JMIR mental health*, 7(6):e19347, 2020.

[36] Koenigbauer Josephine, Letsch Josefine, Doebler Philipp, Ebert David, and Baumeister Harald. Internet-and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. *Journal of affective disorders*, 223:28–40, 2017.

[37] Gabriella Kazai, Paul Thomas, and Nick Craswell. The emotion profile of web search. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1097–1100, 2019.

[38] Carol C Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science*, 42(5):361–371, 1991.

[39] Monica Landoni, Maria Soledad Pera, Emiliana Murgia, and Theo Huibers. Inside out: Exploring the emotional side of search engines in the classroom. In *Proceedings*

of the 28th ACM Conference on User Modeling, Adaptation and Personalization, page 136–144, 2020. URL `https://doi.org/10.1145/3340631.3394847`.

[40] Michael R Liebowitz. Social phobia. *Modern problems of pharmacopsychiatry*, 1987.

[41] Xiaoliang Ling, Weiwei Deng, Chen Gu, Hucheng Zhou, Cui Li, and Feng Sun. Model ensemble for click prediction in bing search ads. In *Proceedings of the 26th International WWW Conference on World Wide Web Companion*, pages 689–698, 2017.

[42] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[43] Daniel Locke, Guido Zuccon, and Harrisen Scells. Automatic query generation from legal texts for case law retrieval. In *Proceedings of the Asia Information Retrieval Symposium*, pages 181–193. Springer, 2017.

[44] David E Losada and Pablo Gamallo. Evaluating and improving lexical resources for detecting signs of depression in text. *Language Resources and Evaluation*, 54(1): 1–24, 2020.

[45] Andrea Sam Lyon and Sheila Mary Mortimer-Jones. The relationship between terminology preferences, empowerment and internalised stigma in mental health. *Issues in Mental Health Nursing*, pages 1–13, 2020.

[46] Thomas F Martinelli, Gert-Jan Meerkerk, Gera E Nagelhout, Evelien PM Brouwers, Jaap van Weeghel, Gerdien Rabbers, and Dike van de Mheen. Language and stigmatization of individuals with mental health problems or substance addiction in the

netherlands: An experimental vignette study. *Health & social care in the community*, 28(5):1504–1513, 2020.

[47] Rishabh Mehrotra, Ashton Anderson, Fernando Diaz, Amit Sharma, Hanna Wallach, and Emine Yilmaz. Auditing search engines for differential satisfaction across demographics. In *Proceedings of the 26th international conference on World Wide Web companion*, pages 626–633, 2017.

[48] Stuart A Montgomery and MARIE Åsberg. A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry*, 134(4):382–389, 1979.

[49] Meredith Ringel Morris, Adam Fourney, Abdullah Ali, and Laura Vonessen. Understanding the needs of searchers with dyslexia. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[50] Yashar Moshfeghi and Joemon M. Jose. On cognition, emotion, and interaction aspects of search tasks with different search intentions. In *Proceedings of the 22nd International Conference on World Wide Web*, page 931–942, 2013. URL `https://doi.org/10.1145/2488388.2488469`.

[51] Zina Moukheiber. Mental health awareness is on the rise, but access to professionals remains dismal. Available at: `forbes.com/sites/zinamoukheiber/2019/05/22/mental-health-awareness-is-on-the-rise-but-access-to-professionals-remains-dismal/#300040c37ba5`, 2019. (accessed March 19, 2020).

[52] Ankit Murarka, Balaji Radhakrishnan, and Sushma Ravichandran. Classification of mental illnesses on social media using roberta. In *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*, pages 59–68, 2021.

[53] Fedelucio Narducci, Marco De Gemmis, and Pasquale Lops. A general architecture for an emotion-aware content-based recommender system. In *Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015*, pages 3–6, 2015.

[54] Michelle G Newman, Andrea R Zuellig, Kevin E Kachin, Michael J Constantino, Amy Przeworski, Thane Erickson, and Laurie Cashman-McGrath. Preliminary reliability and validity of the generalized anxiety disorder questionnaire-iv: A revised self-report diagnostic measure of generalized anxiety disorder. *Behavior Therapy*, 33 (2):215–233, 2002.

[55] National Institute of Mental Health. Mental illness. Available at: `https://www.nimh.nih.gov/health/statistics/mental-illness.shtml`, 2019. (accessed March 19, 2020).

[56] Martin P Paulus and J Yu Angela. Emotion and decision-making: affect-driven belief systems in anxiety and depression. *Trends in Cognitive Sciences*, 16(9):476–483, 2012.

[57] Rebecca M Puhl. What words should we use to talk about weight? a systematic review of quantitative and qualitative studies examining preferences for weight-related terminology. *Obesity Reviews*, 21(6):e13008, 2020.

[58] Chengcheng Qu, Corina Sas, Claudia Daudén Roquet, and Gavin Doherty. Functionality of top-rated mobile apps for depression: Systematic search and evaluation. *JMIR mental health*, 7(1):e15321, 2020.

[59] Esteban A Ríssola, David E Losada, and Fabio Crestani. Discovering latent depression patterns in online social media. In *Proceedings of the 10th Italian Information*

Retrieval Workshop, IIR '19, pages 13–16, 2019. URL `http://ceur-ws.org/Vol-2441/paper10.pdf`.

[60] Esteban A. Ríssola, David E. Losada, and Fabio Crestani. A survey of computational methods for online mental state assessment on social media. *ACM Trans. Comput. Healthcare*, 2(2), March 2021. ISSN 2691-1957. doi: 10.1145/3437259. URL `https://doi.org/10.1145/3437259`.

[61] Saif Saif Mohammad. Word affect intensities. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[62] H Andrew Schwartz, Johannes Eichstaedt, Margaret Kern, Gregory Park, Maarten Sap, David Stillwell, Michal Kosinski, and Lyle Ungar. Towards assessing changes in degree of depression through facebook. In *Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, pages 118–125, 2014.

[63] Sima Sharifirad, Borna Jafarpour, and Stan Matwin. How is your mood when writing sexist tweets? detecting the emotion type and intensity of emotion using natural language processing techniques. *arXiv preprint arXiv:1902.03089*, 2019.

[64] Dushyant Sharma, Rishabh Shukla, Anil Kumar Giri, and Sumit Kumar. A brief review on search engine optimization. In *Proceedings of the 9th International Conference on Cloud Computing, Data Science and Engineering (Confluence)*, pages 687–692. IEEE, 2019.

[65] Judy Hanwen Shen and Frank Rudzicz. Detecting anxiety through reddit. In *Proceedings of the 4th Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*, pages 58–65, 2017.

[66] Anu Shrestha and Francesca Spezzano. Detecting depressed users in online forums. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, page 945–951, 2019. URL `https://doi.org/10.1145/3341161.3343511`.

[67] Robert L Spitzer, Kurt Kroenke, Janet BW Williams, Patient Health Questionnaire Primary Care Study Group, Patient Health Questionnaire Primary Care Study Group, et al. Validation and utility of a self-report version of prime-md: the phq primary care study. *Jama*, 282(18):1737–1744, 1999.

[68] Helen M Startup and Thane M Erickson. The penn state worry questionnaire (pswq). *Worry and its psychological disorders: Theory, assessment and treatment*, pages 101–119, 2006.

[69] Sarah M Tashjian and Adriana Galván. Longitudinal trajectories of post-election distress track changes in neural and psychological functioning. *Journal of cognitive neuroscience*, 32(6):1198–1210, 2020.

[70] Benedikt Till, Marlies Braun, Susanne Gahbauer, Nikolaus Reisinger, Ewald Schwenzner, and Thomas Niederkrotenthaler. Content analysis of suicide-related online portrayals: changes in contents retrieved with search engines in the united states and austria from 2013 to 2018. *Journal of Affective Disorders*, 271:300–309, 2020.

[71] William Tov, Kok Leong Ng, Han Lin, and Lin Qiu. Detecting well-being via computerized content analysis of brief diary entries. *Psychological Assessment*, 25 (4):1069, 2013.

[72] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. Recognizing depression from twitter activity. In *Proceedings of*

the 33rd Annual ACM Conference on Human Factors in Computing Systems, page 3187–3196, 2015. URL `https://doi.org/10.1145/2702123.2702280`.

[73] Frans van der Sluis and Betsy van Dijk. A closer look at children's information retrieval usage. In *33st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'10)*, 2010.

[74] Nikhita Vedula and Srinivasan Parthasarathy. Emotional and linguistic cues of depression from social media. In *Proceedings of the 2017 International Conference on Digital Health*, page 127–136, 2017. URL `https://doi.org/10.1145/3079452.3079465`.

[75] Kirsten G Volz and Ralph Hertwig. Emotions and decisions: Beyond conceptual vagueness and the rationality muddle. *Perspectives on Psychological Science*, 11(1): 101–116, 2016.

[76] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1), March 2018. URL `https://doi.org/10.1145/3191775`.

[77] Luyan Xu, Xuan Zhou, and Ujwal Gadiraju. Revealing the role of user moods in struggling search tasks. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 1249–1252, 2019. URL `https://doi.org/10.1145/3331184.3331353`.

[78] Borchuluun Yadamsuren and Jannica Heinström. Emotional reactions to incidental exposure to online news. *Information Rsearch*, 16(3):16–3, 2011.

[79] Yahoo! Yahoo! datasets. Available at: `https://webscope.sandbox.yahoo.com/catalog.php?datatype=l`, 2019. (accessed November 6, 2019).

[80] Anis Zaman, Rupam Acharyya, Henry Kautz, and Vincent Silenzio. Detecting low self-esteem in youths from web search data. In *The World Wide Web Conference*, page 2270–2280, 2019. URL `https://doi.org/10.1145/3308558.3313557`.

[81] Haoran Zhang, Amy X Lu, Mohamed Abdalla, Matthew McDermott, and Marzyeh Ghassemi. Hurtful words: quantifying biases in clinical contextual word embeddings. In *proceedings of the ACM Conference on Health, Inference, and Learning*, pages 110–120, 2020.

[82] Junqi Zhang, Yiqun Liu, Shaoping Ma, and Qi Tian. Relevance estimation with multiple information sources on search engine result pages. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 627–636, 2018.

[83] Changye Zhu, Baobin Li, Ang Li, and Tingshao Zhu. Predicting depression from internet behaviors by time-frequency features. In *Proceedings of the 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 383–390. IEEE, 2016.