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

ASHLEE MILTON*, GroupLens - Dept. of Computer Science, University of Minnesota, USA
MARIA SOLEDAD PERA†, Web Information Systems - TU Delft, The Netherlands

Researchers worldwide have explored the behavioral nuances that emerge from interactions of individuals afflicted by mental health disorders (MHD) 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: web search engines (SE). Each day, users with MHD embark on information seeking journeys using popular SE, like Google or Bing. Every step of the search process for better or worse has the potential to influence a searcher's mindset. In this work, we empirically investigate what subliminal stimulus SE present to these vulnerable individuals during their searches. For this, we use synthetic queries to produce associated query suggestions and search engine results pages. Then, we infer the subliminal stimulus present in text from SE, i.e., query suggestions, snippets, and web resources. Findings from our empirical analysis reveal that the subliminal stimulus displayed by SE at different stages of the information seeking process differ between MHD searchers and our control group comprised of "average" SE users. 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.

CCS Concepts: • **Social and professional topics** → *User characteristics*; • **Information systems** → **Sentiment analysis**; *Web searching and information discovery*.

Additional Key Words and Phrases: Mental Health, Web Search Engines, Sentiment, Emotion

ACM Reference Format:
Ashlee Milton and Maria Soledad Pera. 2023. Into the Unknown: Exploration of Search Engines' Responses to Users with Depression and Anxiety. 1, 1 (March 2023), 29 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Persuasive technologies, embedded into almost every facet of our lives, are designed to influence the behaviors or attitudes of individuals [28]. Nevertheless, not everyone is swayed by persuasive technologies in the same manner. Consider individuals suffering from mental health disorders (**MHD**), such as schizophrenia or bipolar disorder. Their rational decision-making is inhibited by their emotional states [85], causing them to be more susceptible to persuasion than individuals not ailed with a MHD. This seems to imply that the manner in which MHD users would interact with and be affected by persuasive technologies differently. With mental illness being a persistent issue [2]–as per the National Institute of Mental Health, one in five adults in the United States suffers from

*Work conducted while a MS student at Boise State University
†Work partially conducted while still affiliated with Boise State University

Authors' addresses: Ashlee Milton, GroupLens - Dept. of Computer Science, University of Minnesota, Minneapolis, USA, 55455, milto064@umn.edu; Maria Soledad Pera, Web Information Systems - TU Delft, Delft, The Netherlands, 2628 XE, m.s.pera@tudelft.nl.

XXXX-XXXX/2023/3-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

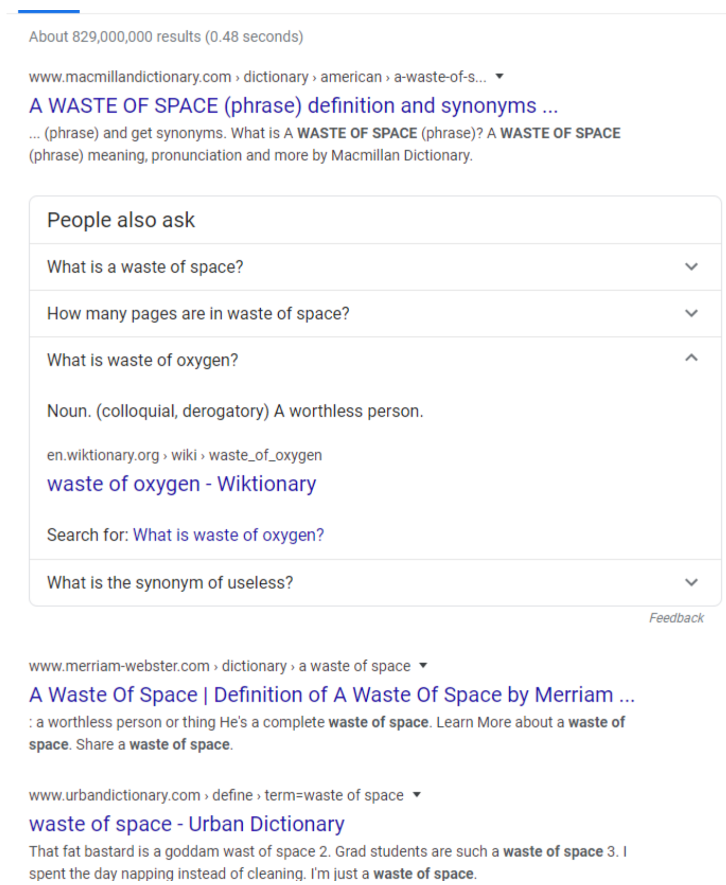

Fig. 1. Snapshot of Google's results retrieved for the query "waste of space." (retrieved June 2021)

a MHD [63]–it is imperative to identify and understand the consequences that might unknowingly occur when people suffering from MHD engage with persuasive technologies.

Web search engines (**SE**) are a ubiquitous persuasive technology, yet there is little information regarding how MHD users engage with them. See a snapshot of the results retrieved by Google for the query "waste of space", a phrase many people with depression may state or think, in Figure 1. The snippets include 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 be blindsided by and feel distressed by such phrases appearing in response to their search. 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 the information seeking process (**ISP**), as defined by Kuhlthau [41], where users engage with SE. 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 align 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. To characterize the users who prompt these algorithmic responses, we examine the affect and MHD indicators present in users **queries**–initiation. 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 scrutinize the affect and MHD indicators inferred from queries and subsequent SE responses to traditional searchers, which we treat as our **control** group.

To manage scope, we center our study on *adults* who perform online inquiry tasks in *English* using popular *commercial SE* (Google and Bing), as they are mainstream among English speakers. In the context of this work, we assume searchers to be presumably based in the United States, as data collection took place with in a United States settings. Of note, a large portion of United States adults suffers from a MHD (46.6 million in 2017 [63]). Further, it is well-documented that the degree to which subliminal language influences users with MHD depends on the kind of MHD they have [6, 51, 54, 66]. To not overgeneralize, we only examine *depression* and/or *anxiety*[2].

With this work, we do not study MHD users themselves, i.e., how they perceive and react to stimulus inadvertently produced by SE. Instead, we focus on algorithmic functionality and aim to establish a foundation on how SE respond on a subliminal level to users with MHD during their information seeking journey. To guide our exploration, we define two research questions:

RQ1 What subliminal stimulus do SE responses project directly onto users with MHD?
RQ2 How do subliminal stimulus of SE responses indirectly change through the ISP for MHD users?

For analysis purposes, access to **text samples** that capture ISP interactions that represent MHD and control group searchers, i.e., query suggestions, snippets, and web resources, is vital. Unfortunately, large-scale query logs from commercial 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. With these corpora, we generate datasets that imitate the 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 [49, 59, 72].

The contributions of this work include: (i) generating *subliminal profiles* of SE responding to interactions with MHD users, (ii) creating three domain-specific lexicons, based on social media posts as well as psychological surveys for depression and anxiety, and (iii) offering an in-depth look at the current state of SE responses to MHD searchers through the comparison of system-user interactions generated by MHD and control users while highlighting limitations of popular SE and implications informing future research. The presented analysis could serve 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. Implications of this work could extend into a variety of areas, e.g., examining whether lessons learned from this work apply internationally, as MHD present themselves and are dealt with differently among varying languages and cultures.

[1] We use the psychology view of *affect*: “any experience of feeling or emotion” [4].
[2] From here on, whenever we state MHD, we refer to depression and/or anxiety.

Mental illnesses are also prominent among younger populations, particularly teens with depression and anxiety. Thus, our investigation can help frame further inquiries into the problems when focused on this audience.

2 RELATED WORK

MHD are a prevalent concern, one that has received attention from researchers and practitioners alike [20, 38, 67, 86]. This is also evident in the discussions and works presented at ongoing workshops, such as eRisk and CLPsych, at prominent venues, like CLEF and ACL [22, 53, 64, 94]. The common denominator among existing literature is that it is based on how people with MHD turn to technology or how interactions with technology can be used to reveal MHD-related aspects (e.g., depression detection on social media [45, 71]), but not how technology influences them, which is the focal point of this work. In the rest of this section, we discuss the background and related literature that contextualize our research endeavors.

2.1 MHD and Social Media

From a persuasive technology perspective, MHD literature centers on the social media domain. Mainly, the ability to identify users with depression from social media posts [18, 24, 25, 69, 73, 82]. Depressed users have not been the only ones considered. Researchers have also studied the linguistic qualities of social media posts by users with other MHD, such as schizophrenia [15], depression, anorexia, self-harm, and post-traumatic stress disorder [68]. Findings from research conducted in previous literature 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., the number of posts and retweets) [77, 84], as well as the trends in interactions of MHD users with other social media users (e.g., who users talk to and how often) [68, 84]. In an attempt to detect the state of mind of social media users, several researchers have also examined the tendencies 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 byRíssola et al. [70], 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 work spotlights SE.

2.2 MHD and SE Interactions

Research exploring the relationship between MHD and SE is in its infancy. From a user perspective, Campbell et al. [17] discuss the help-seeking behavior of users with MHD, i.e., searchers looking for resources to understand and help with MHD. Zhu et al. [93], on the other hand, use query logs from a university webserver to predict users suffering from depression. Similarly, Zaman et al. [90] identify searchers with self-esteem issues from user-provided Google search histories. Birnbaum et al. [15] have contemplated the feasibility of detecting the early onset of MHD from query logs. Instead, Xu et al. [87] turn to query logs to evaluate the degree to which mood influences users' interactions with SE. While not focused on MHD, Moshfeghi and Jose [58] bring up a 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 depending on the search task or the specific intentions of a user when conducting a search, different emotions can be experienced at varying levels. Ever since March of 2020, the world has been amid a global pandemic (i.e., COVID-19). This has prompted researchers to study if and how search trends for mental health have changed [7, 36]. Outcomes reveal that mental health queries are more prevalent now, evidencing the need for explorations such as the one we present in this work.

2.3 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 do not consider emotions. Azpiazu et al. [8] and Locke et al. [48] 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. [92] utilize visual aspects of SERP to estimate the relevance of a resource, whereas Ling et al. [46] use ensemble models to predict ad click-through rates on SERP. The works of Gossen [29] and Morris et al. [57] 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. [80], 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. [39] and Demartini and Siersdorfer [26] investigate the emotions, sentiments, and opinions emerging from web resources, yet only in response to queries formulated by traditional users. Additionally, Landoni et al. [42] explore SERP and emotions, but in their case the population under study are children.

In short, previous works on MHD and SE have investigated user interactions from the perspective of help-seeking, self-esteem, and mood. However, none study the search systems themselves to ascertain the potential alterations SE responses have on the mental well-being of users with MHD. Although the works we discuss in this section highlight SE functionality or affective responses at each ISP stage, none directly studies the comprehensive subliminal stimulus responses of different SE functionality for MHD individuals. 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.

3 EXPERIMENTAL SETUP

Here, we describe the experimental setup of the study conducted to answer our research questions.

3.1 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 while synthetic still exemplifies MHD searchers' interactions with SE. Reddit is the go-to source of mental health data for research in the social media space and has been for many years [23, 25, 50, 61, 68]. This prompted us to turn to Reddit for our exploration.

To transform Reddit posts into synthetic queries, we emulate the text processing steps common in social media explorations pertaining to mental health which often depend on examining terms and phrases in the quest for terms related to MHD.[3] In particular, we use the framework outlined by De Choudhury et al. [24] to isolate from social media data phrases common among individuals affected by depression. In our case, we instead recognize phrases containing terminology potentially used by MHD searchers that can serve as a proxy for search queries. For this, we extract from Reddit posts related to the topics of MHD and use phrases extracted from these posts as our queries. 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. Specifically, we use 2,600 posts extracted

[3]Note that much like in similar works in the social media realm, we do not perform verification of users' diagnosis with depression or anxiety. We simply try to emulate the vocabulary that MHD searchers could be used to initiate an online inquiry, as opposed to determining their diagnostic status.

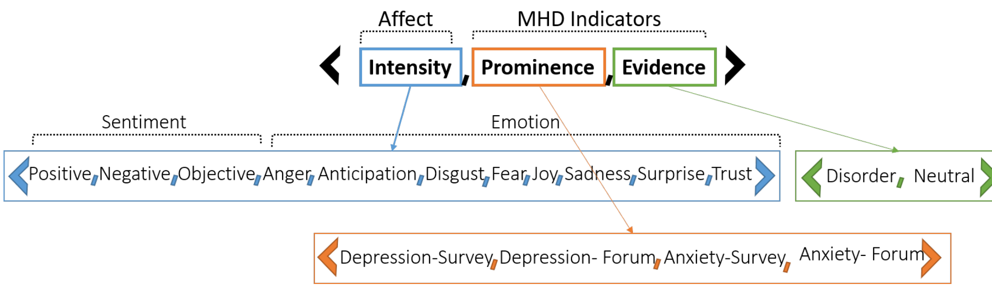

Fig. 2. Structure of the stimulus vector representing each text sample.

from 10 subreddits: *r/AnxietyDepression*, *r/Anxiety*, *r/Depression*, *r/DepressionHelp*, *r/MentalHealth*, *r/SocialAnxiety*, *r/OffMyChest*, *r/SelfHarm*, *r/MentalHealth*, and *r/SuicideWatch*. This results in 4,418 synthetic queries: 1,200 unigrams and the remaining n-grams, mimicking the distribution of unigrams and n-grams of our control sample discussed later in this section.

We use these generated queries to mimic users' interactions with SE and elicit responses, resulting in text samples representing QS, SERP[4], and RR for Google[5] and Bing[6] using their respective API.[7] For data gathering purposes, we only record the first SERP, as users do not often go past the first page when looking at search results [75]. Further, for each web resource on a SERP, we extract its title, snippet, and full web content. 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.

As part of our analysis, we also examine interactions from traditional users, who we treat as a control group. Thus, we need data capturing their interactions. In this case, we use a sample of 4,458 queries made available for research purposes by Yahoo Webscope's [89]: 1,211 unigrams and the remaining n-grams. We use these queries to generate the appropriate text samples, following the same data gathering procedure outlined for MHD searchers.

3.2 Establishing Stimulus Vectors

To characterize the subliminal stimulus expressed by SE, we explore the text samples (e.g., query suggestions, resource titles, snippets, and web page content) gathered for both MHD and control group users 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*, illustrated in Figure 2, which serves as a representation of the affect and MHD indicators and accounts for each aforementioned perspective in the corresponding text sample.

3.2.1 Intensity of Affect. Lexicons are a starting point for discerning affective language in text samples; they are commonly used in the social media domain for analysis of MHD [24]. The ones most relevant to our study are those specific to sentiment and emotion. Following the framework presented by Kazai et al. [39], which represents the sentiments and emotions emerging from text samples as a distribution of intensity scores, we create an affect **intensity** vector depicting the affect distribution of text samples. For **sentiment**, we use SentiWordNet [9]. In this lexicon, sentiment is depicted as `Positive`, `Negative`, and `Objective`. For **emotion** identification, we use Emotion Intensity Lexicon (NRC-EIL) [72], a lexicon that characterizes words as vectors of `Anger`, `Anticipation`,

[4]We treat title and snippets on SERP as snippets and refer to them as such
[5]https://developers.google.com/custom-search
[6]https://www.microsoft.com/en-us/bing/apis/bing-web-search-api
[7]Both APIs were set up to not collect personal data. Further, data collections was done from the same IP and location.

`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 affective scores for each of its words.

3.2.2 Prominence of MHD. The terminology that MHD individuals both use and respond to is noticeably different than that of traditional users [13]. Thus, we 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 899 terms described in [49] into a single lexicon we refer to as `Depression-Forum`.

To our knowledge, there are no domain-specific lexicons for anxiety. Consequently, we adopt the procedure outlined by De Choudhury et al. [24] for building a depression lexicon, 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 anxiety. In our case, we use the Reddit data collected by the authors of [76] 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 overlap with those identified to be part of the lexicon, we use a subset of 1.6 million Wikipedia articles (unlike the full Wikipedia originally used in [24]). Further, given our interest in terms related to anxiety, we use the regex "anx*" for bigram generation. This results in our `Anxiety-Forum` lexicon, which contains 79 terms.

Our depression and anxiety lexicons are constructed from social media posts, and thus may overlook formal terms that psychologists would consider in MHD diagnosis. With that in mind, we construct two new lexicons, one for depression and another for anxiety, comprised of terminology inferred from several psychological assessments: (1) State Worry Questionnaire (PSWQ) [79] (2) Liebowitz Social Anxiety Scale (LSAS-SR) [44] (3) Hamilton Anxiety Rating Scale (HAM-A) [32] (4) Anxiety Symptoms Questionnaire (ASQ) [10] (5) Generalized Anxiety Disorder (GAD) Screening Tool [62] (6) Beck's Depression Inventory [11] (7) Patient Health Questionnaire-9 (PHQ-9) [78] (8) Hamilton Depression Rating Scale (HAM-D) [33] (9) Montgomery and Åsberg (MADRS) Depression Rating Scale [56] (10) EQ-5D [35] (11) Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [5].

These assessments represent a range 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. While these assessments vary greatly in age, all are still used by mental health professionals for the assessment of patients [43]. 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 what a statement portrays. As these statements contain language meant to resonate with people with MHD, we depend upon an impartial assessor to identify and extract the 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.

3.2.3 Evidence of MHD. Contemplate the statement "I will never be happy again". From the individual keywords, one could assess it to be "happy" in tone. Yet, associating "never" to other keywords in the phrase reveals the real tone of the phrase: sadness. Further, "I" is prefacing a negative emotion. This is a linguistic style commonly seen among depressed individuals who are known to frequently use self-referencing statements with negative emotions. When examining terms in isolation, it is possible to miss nuances emerging by considering 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. [59]. This strategy utilizes a RoBERTa model [47] that explores a text (specifically the 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 [59] 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 we augment original text processing by removing special characters, expanding contractions, and correcting misspelled words. We train the adapted strategy using the same libraries, parameters, and data as in [59].[8] Using the trained model, we create an MHD **Evidence** vector for each text sample. This vector captures the probability distribution for each class, `Disorder` or `Neutral`, in the range of 0 to 100; the closest to 100 the more indicative of the respective class.

3.3 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. Thus, we combine the aforementioned vectors for a given ISP stage into a single *subliminal stimulus profile* which serves as a snapshot of the affect and MHD indicators presented to searchers as a result of their interactions with the SE.

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 over several queries when SE respond to users, 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 whether 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 [37]. This leads us to create a *rank-1 profile*. In this case, we average the first ranked text sample vector for each query used to generate SE results.

Example QS Query Log:	'*Apple*': ['*Apple watch*', '*Apple crisp*', '*Apple picking*'] '**Orange**': ['**Orange aesthetic**', '**Orange juice**', '**Orange chicken**']
Overall Profile:	('*Apple watch*' + '*Apple crisp*' + '*Apple picking*' + '**Orange aesthetic**' + '**Orange juice**' + '**Orange chicken**') / 6
Query-Based Profile:	(('*Apple watch*' + '*Apple crisp*' + '*Apple picking*') / 3 + ('**Orange aesthetic**' + '**Orange juice**' + '**Orange chicken**') / 3)) / 2
Rank-1 Profile:	('*Apple watch*' + '**Orange aesthetic**') / 2

Fig. 3. Representation of the aggregation strategies for generating subliminal stimulus profiles. The green and italicized QS are in response to 'Apple' and the orange and bold QS are in response to 'Orange'.

Example. To illustrate the generation of each profile, suppose we have data containing the text samples for the queries'apple' and 'orange' and we are considering SE functionality QS. A visual

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

representation of this example can be seen in Figure 3. 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-1 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.

3.4 Generating Datasets and Associated Subliminal Stimulus Profiles

We use the text samples collected in Section 3.1 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 (Section 4). 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 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 Section 3.3. 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.

3.5 Tests of Statistical Significance

To understand the stimulus SE convey to users with MHD, we compare and contrast in Section 4 the various pairs of subliminal profiles described in Section 3.4, 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 $\alpha = 0.05$; $N = 15$, which is the number of vector components) to indicate changes in stimulus across pairs of profiles that are statistically significant.

4 EXPERIMENTS AND ANALYSIS

In order to answer our research questions, we conduct in-depth analysis and share subsequent findings from our experiments. We explore SE responses at each of the main ISP stages, as well as the stability in stimuli across different stages of the information seeking journey, e.g. queries to query suggestions. When pertinent, we compare the subliminal stimulus of SE responses to MHD searchers with those presented to the control group, offering further insights into the affect and MHD indicators MHD searchers are subjected to.

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

To depict the affect SE project to MHD users, who are known to be easily influenced by external stimuli [5], we explore what subliminal stimuli SE directly convey to MHD searchers through **QS**, **SERP**, and **RR**. 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 profiles introduced in Section 3.4 (excluding query profiles), which we examine from diverse perspectives. We showcase trends; that is fluctuations in the stimulus profiles across affect and MHD indicators. observed when aggregating stimulus profiles of text samples *overall* and *rank-1* on each ISP stage in Figure 4.

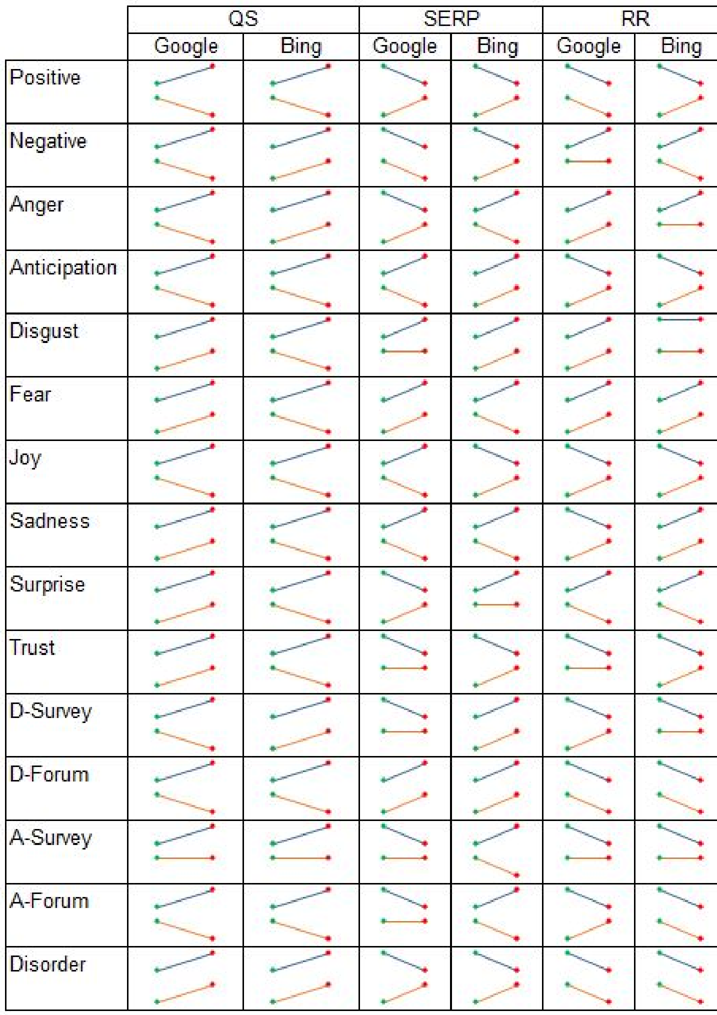

Fig. 4. Subliminal stimulus trends between overall and rank-1 aggregation profiles for QS, SERP, and RR for Google and Bing. Trends lines in blue (top) refer to MHD searchers, orange (bottom) is the control group. Green points (left) refer to overall profiles, red (right) is rank-1.

4.1.1 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 by offering other queries and thus different information [27]. We first dissect Google's QS profiles in Table 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 stood out: `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 among these QS samples the presence of 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” 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 the **Intensity** vectors is the culprit for the high scores for upbeat emotions and therefore these scores may misrepresent 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 terminology is the most prominent. 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 the QS profiles, `Disorder` is lower than `Neutral`, hinting at writing patterns associated with MHD users not being prominent among QS.

Table 1. Subliminal stimuli profiles of QS generated by Google, blue indicates significant differences between MHD profiles and their corresponding control counterparts ($p < 0.01$); bold ($p < 0.05$).

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-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 their similitude, there are some peculiarities distinguishing profile score distributions across aggregation strategies. The profile capturing stimuli of the top QS (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 aggregations 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 juxtaposing the **Evidence** vectors 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 is clear that Google’s QS for MHD searchers contain more MHD indicators. 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, i.e., there are less dramatic spikes in the control stimulus scores than those observed in counterpart MHD profiles (see Figure 5). Comparing the **Intensity** vectors for 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 by analyzing the profiles from 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 argument on the discrepancy being due to the word-independence assumption of the approach for intensity estimation.

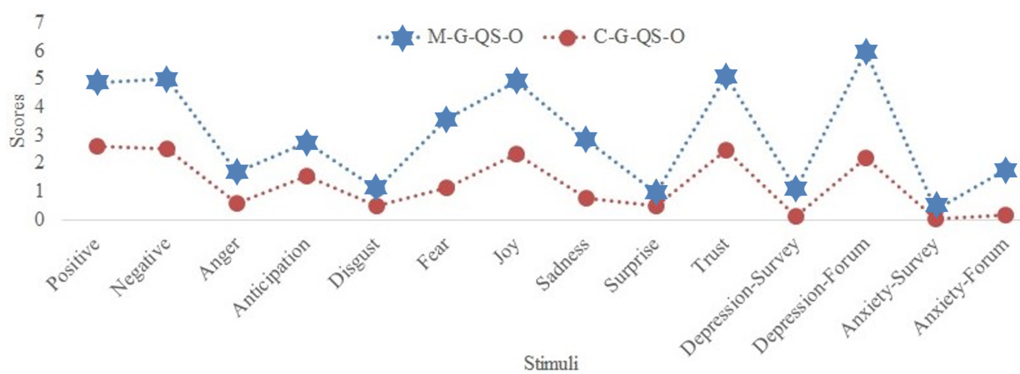

Fig. 5. Representation of stimuli for M-G-QS-O and C-G-QS-O. For illustration purposes, we omit `Objective`, `Disorder`, and `Neutral` from the corresponding subliminal stimuli profile representation. It emerges from this image that while both profiles have similar distribution scores, there are visible spikes in the intensity of *Negative*, `Anticipation`, *Fear*, `Joy`, *Sadness*, `Trust`, `Negative`, `Depression-Forum`, and *Anxiety-Forum*.

Table 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*			
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum	Disorder	Neutral
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 2. We observe a significant 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 increases. 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. To bring into perspective the diverse stimulus MHD users experience, we juxtapose the profiles of QS presented to MHD searchers versus those shown to the control group. There is a proportionally large disparity between the stimulus scores in the **Intensity** vector in the control group profiles and those on MHD profiles. `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 respective control counterparts are statistically significant.

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 for Bing, 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 the counterpart control profiles. QS contain more terminology related MHD for MHD users, but not as much as Google's QS. 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 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 mechanisms that may be unhealthy for said users [30, 40].

4.1.2 The Hidden Messages of SERP. We shift our attention to the subliminal stimuli conveyed by SERP for Google and Bing. Focusing on Google's SERP (Table 3), when compared with M-G-SERP-O and M-G-SERP-BQ, we observe in M-G-SERP-R1 a significant decrease in `Disorder`, as well as an increase in `Neutral`. While not statistically significant, changes in `Anticipation` and `Disgust` have noticeable increases for rank-1 profiles. The high scores for `Disgust` and `Anticipation` on snippets positioned at the top of SERP, could be a concern as the first thing MHD users see on a SERP, `Disgust` and `Anticipation` could be a catalyst for unexpected behavior [27]. This is 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 note that the majority discuss topics about sickness, sufferings, illness, stigmas, 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 [52]. Examining MHD profiles for Bing's SERP in Table 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 `Depression-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 from SERP in M-B-SERP-R1. We find that snippets tend to address topics like anxiousness, decision-making, and nervousness. These topics align with anxiety, implying that Bing could perpetuate 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 responding to MHD users have proportionally higher scores with SERP produced by Bing than Google, leading 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

Table 3. Subliminal stimuli profiles of SERP generated by Google, blue indicates significant differences between MHD profiles and their corresponding control counterparts ($p < 0.01$); bold ($p < 0.05$).

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
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 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*			
	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
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

higher `Positive` scores than `Negative` scores which we did not expect given 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 4.1.1. Consequently, we cannot attribute the elevated `Joy` and `Trust`, which are present in the MHD SERP 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 4.1.1. For example, in Table 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 4.1.1. The subliminal stimuli for MHD and traditional searchers are more similar for SERP than they are for QS.

Examining the gap between the stimulus in the SERP profiles of MHD and traditional users for both SE, it seems that SERP are less stimulating than QS regardless of SE for both types of users. The emotions `Anger`, `Fear`, and `Sadness` 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 exposed to less than pleasant emotions just not as predominately as they would when interacting with QS.

4.1.3 The Essence of Retrieved Resources. Nearing the end of the information seeking journey, searchers read through the content of clicked results. Retrieved resources result in longer text samples than QS or snippets. Consequently, we would expect them to contain a broader range of stimulus. We first examine resources retrieved by Google. We inspect 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). 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` have significant decreases between the profiles averaged by rank-1 (M-G-RR-R1) and by-query (M-G-RR-BQ). When contrasting the MHD user and control profiles for RR all observed differences in stimulus scores are significant.

In the juxtaposition of Bing's profiles of RR for MHD and traditional searchers in Table 6, we note that, excluding the difference in `Joy` between M-B-RR-R1 and C-B-RR-R1 which are not significant, all other divergences 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 trends in the stimuli of RR, Bing has higher stimulus levels and a spike in `Anticipation` that it is not visible among Google's RR. Neither Google's nor Bing's first-ranked RR have higher scores in stimulus than those reported among top-10 RR, a departure from what we saw for QS and snippets. Overall, in spite of the few reported variations of stimuli, for the most part, RR follow the same trends as those seen for QS and SERP in Sections 4.1.1 and 4.1.2, respectively.

Table 5. Subliminal stimuli profiles of RR generated by Google, blue indicates significant differences between MHD profiles and their corresponding control counterparts ($p < 0.01$); bold ($p < 0.05$).

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-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 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*			
	Pos	Neg	Obj	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Survey	Forum	Survey	Forum	Disorder	Neutral
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

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

From a search query that initiates the search process it is possible to infer the affect and any MHD indicators that users disclose to a SE via the language they use when expressing their information needs [74, 81]. SE are an outside source of stimuli, which MHD users are known to be susceptible to [5, 30, 40]. We posit that 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 [27, 65, 88]. In turn, this could inadvertently alter these users' information seeking journeys. Given the opportunities that SE have to influence users throughout the ISP, we need to fully understand the subliminal stimuli of SE responses and their potential repercussion on searchers. Thus, we scrutinize if and how stimuli from SE responses fluctuate in affective tone and MHD language usage.

Consider a user seeking for an entertainer for a child's birthday party using the query "clowns". In response to this query, Google produces QS ranging from "clowns scary" to "killer klowns from outer space" to "clowns for hire". The query itself could be perceived as "happy" whereas the QS have an overall feeling of "fear". Further, the first SERP result 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 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 resource contains a picture of a clown holding a machete and words like "panic", "threats", and "creepiness". What started as a benign task quickly turned horrifying in a few clicks. 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 3.4. To refer to these profiles, we use a similar naming convention as the one introduced in Section 3.3 for profiles of SE responses, but with only three letters: **{M, C}-{Q}-{O}**. We then contrast the profiles of adjacent ISP stages to observe fluctuations that can occur from the initiation of a search to the presentation of RR. Specifically, we compare (i) the stimulus profiles of users' queries with respect to the corresponding profiles originating 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. Much like we did in our empirical analysis for RQ1 (in Section 4.1), 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. In the analysis reported in Section 4.1 we did not see many variations among the stimuli of profiles using different aggregations strategies. Consequently, we only focus on the profiles generated using *overall* as the aggregation strategy. Whenever merited, we do point out notable developments resulting from profiles generated using either *by-query* or *rank-1* aggregation strategies. 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 6.

4.2.1 From Queries to Query Suggestions. To establish the stimulus manifested from users' queries, we turn to the query profiles in Table 7. 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. 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 for 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 differences in MHD indicators, it is visible that MHD users start their information seeking journey in a very different state of mind than traditional users.

To gauge whether the stimulus from QS produced by SE diverge from the stimulus conveyed in users' queries, we examine the profiles for users' queries vs. those for associated QS generated by Google or Bing. Starting with Google, from M-Q-O and M-G-QS-O in Table 8 we observe that differences in scores in the profiles for QS vs Q are statistically significant except for `Anxiety-Survey`

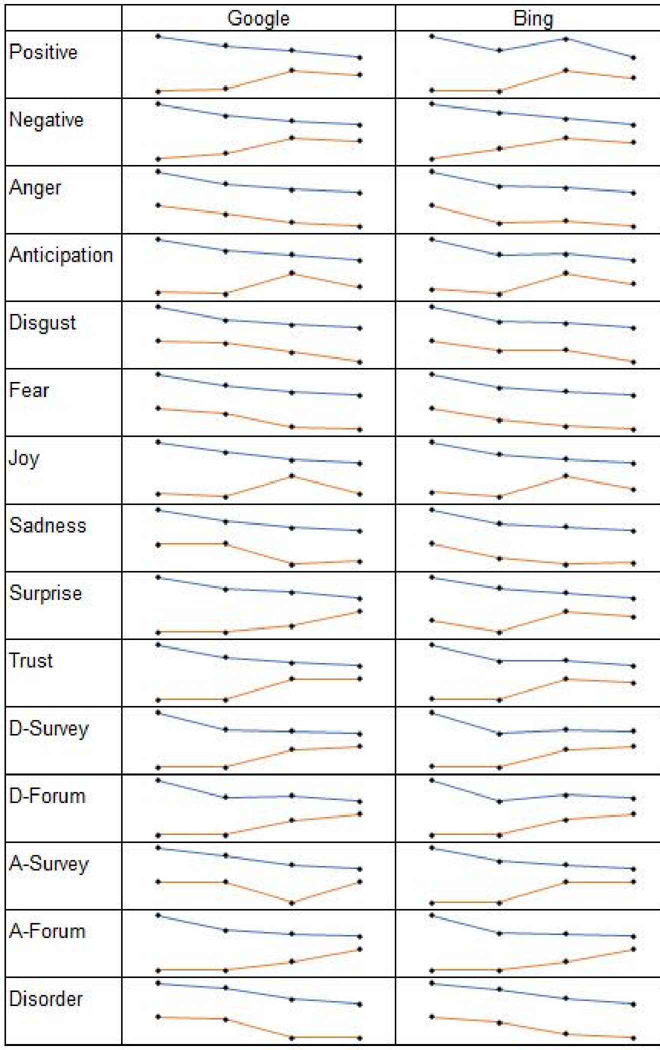

Fig. 6. Subliminal stimulus trends across the ISP for Google and Bing. Blue trend lines (top) refer to MHD searchers, orange for the control group (bottom). Points along trend lines represent from left to right Q, QS, SERP, and RR.

Table 7. Subliminal stimuli profiles of Q, blue indicates significant differences across MHD and control profiles ($p < 0.01$).

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
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

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 has 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 user's query. This shows that while matching the stimuli of the queries of MHD users, Google mutes the stimulus potency.

Table 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 transitions among traditional users. We explore distribution trends between C-Q-O and C-G-QS-O (i.e., profiles for the control group) seeking to confirm if they remain the same as the ones detected for MHD searchers. From Table 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 than 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 or simply no changes 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 shift in stimuli from Q to QS in Google are not typical as there are differences in the stimuli SE response convey to MHD versus control users, with the affect from MHD queries deadening in the stimulus QS profile.

By comparing M-Q-O and M-B-QS-O profiles (in Table 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. We 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. There are significant changes between C-Q-O and C-B-QS-O for the **Evidence** vector. The differences noted between traditional Q and QS on Bing, 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 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 the offered suggestions 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 [60]). We surmise that by dulling the stimulus in QS the SE could keep MHD users from finding support (e.g., song lyrics they find comforting) or instead lead them stray towards results that contain triggers for their MHD.

Table 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 ($p < 0.01$); bold ($p < 0.05$).

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-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

4.2.2 From Queries to SERP. We next peruse the transition from queries to SERP by comparing M-Q-O to M-G-SERP-O (Table 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 diverges from the stimulus of corresponding search queries. When considering Bing's SERP generated from users' queries (M-B-SERP-O c.f. C-B-SERP-O in Table 11), emerging trends for the most part closely resemble Google's. 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 display similar stimuli to the MHD users' queries prompting the SERP, but the strength with which they appear is greatly diminished from the users' original affect, which causes a disparity between users' mental states and the SE responses.

Table 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*			
	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-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 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 ($p < 0.01$); bold ($p < 0.05$).

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-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 study the stimuli of the SERP presented 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 variations we detect between M-Q-O to M-G-SERP-O in the number of stimuli that had significant changes. The changes to stimuli for MHD searches mostly decrease in stimulus scores, while for traditional searchers bleak emotions decrease and upbeat ones increase. When exploring Bing's responses to the control group we see that with the exception of `Disgust`, `Surprise`, and `Anxiety-Survey`, all modifications in stimulus are statistically significant for C-Q-O vs. C-B-SERP-O. Additionally, we see the same elevation of upbeat and decrease of bleak emotions in Bing that was present in Google.

From findings arising as a result of analyzing transitions in stimulus from queries to SERP on MHD and control profiles, we deduce that both SE produce SERP with dissimilar stimulus from queries for both MHD and traditional searchers. However, 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 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.

4.2.3 From SERP to RR. As previously stated, the subliminal stimulus of QS and SERP are known to mutate the affect and MHD indicators users express in their queries. We have not considered, however, how SE express the stimuli of web content through SERP. To investigate potential variations in stimuli from SERP snippets 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 12). except for `Objective` and `Neutral`, we see significant decreases in stimulus between SERP and RR for MHD users. When considering Bing (M-B-SERP-O vs. M-B-RR-O in Table 13), we see a very similar trend as 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 score for `Positive` between RR and SERP than Google does. The gap in stimulus between SERP and RR implies that regardless of the SE, SERP amplify the stimulus arising from RR for MHD users, which can be problematic when MHD users click on resources on emotional subjects as there is a gap between the expectation and reality in terms of stimuli of results.

Table 12. Subliminal stimuli profiles of SERP and RR generated by Google, blue indicates significant differences between SERP and RR ($p < 0.01$); bold 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 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

Consider C-G-SERP-O and C-G-RR-O. It is evident that there are more stimuli for which variations are not significant (`Fear`, `Sadness`, `Trust`, `Anxiety-Survey`, and `Disorder`) when comparing with

MHD users profiles. Additionally, the magnitude of changes between SERP and RR profiles are smaller for traditional than for MHD users. When examining C-B-SERP-O and C-B-RR-O we see fewer variations, 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 vary in how many scores have significant changes, with overall having the most changes and by-query having the least. largely, there are fewer discrepancies between RR and their SERP representations for traditional users than there are for MHD users; in general, SERP are more representative of RR 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" (related to 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 which 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 case, SE inadvertently emulate the actions of clickbait 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 needs.

4.3 Discussion

Outcomes from our analysis reveal several interesting phenomena that occur when MHD searchers interact with SE (see illustrative trend highlights in Figures 4 and 6).

We first examined each of the ISP stages as snapshots (RQ1), evidencing that changes to stimulus responses do transpire at all stages of the ISP. Further, studied SE respond to users with MHD with elevated levels of `Anger`, `Fear`, and `Sadness` at all ISP stages, a departure from the changes seen with the control users. These findings are concerning, as prolonged exposure to such cynical emotions can be damaging to the mental health of users with MHD [34]. Additionally, traditional searchers are not exposed to elevated levels of bleak emotions (`Anger`, `Disgust`, `Fear`, and `Sadness`); they tend to be presented with more neutral stimulus, indicating that SE responses differ depending on user type. How exaggerated cynical emotions become for MHD users varies across ISP stages; most prominent in QS and the least in RR. The ranking position of a SE response impacts the potency of stimuli conveyed to users–top-ranked responses consistently display slightly more intense emotions than those observed when averaging the intensity of the emotions of all alternatives considered for QS, SERP, or RR. MHD users are then faced with cynical emotions on the first response they see. This is something that cannot be overlooked if we consider that users will scan through top results (snippets) to determine their relevance, yet being forced to read such snippets could be harmful to users with MHD. While investigating into top-ranked responses, we noticed that they could boost detrimental combinations of emotions, recall our dive into `Anticipation` and `Disgust` in Section 4.1. We argue that this implies 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 causing unneeded strain on the already compromised mental health of MHD searchers. Most of the observed trends apply to both Google and Bing; the latter often features higher stimulus scores.

Information seeking stages are not isolated, they are part of a process. This impelled us to turn our attention to how stimuli explicitly differ from one ISP stage to the next (RQ2). The emotions and MHD indicators SE express throughout the ISP can influence the way users internalize information, making it vital that we understand how the stimulus SE convey is altered from one stage to the next.

From our analysis of results related to RQ2 explorations, we see that stimulus gets progressively more stable throughout the ISP, with RR having the most stable profiles among those generated for different types of SE responses. Additionally, RR have the least differences between subliminal stimulus profiles of traditional and MHD users when compared to QS and SERP. Still, users with MHD are presented with proportionally higher cynical emotions than the control counterparts in RR. Moreover, it emerges that the affect in a user's original query gets distorted gradually over the ISP. The stimuli inferred for traditional users' queries are stabilized through SE responses, but for MHD users their expressed affect and MHD indicators are instead dampened. To illustrate this phenomenon, visualize a funhouse mirror that reflexs back what is in front of it, returning a distorted vision of what it saw. In response to their queries, traditional users receive a more stable version of what they started with. Unfortunately, MHD users see an image with less contrast, the dark is 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.

Insights gained from analysis associated with RQ1 and RQ2 have enabled us to describe 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?*. We encapsulate the direct possible implications in a real-world scenario, with the following example:

> *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 portray high levels of fear and sadness, as well as 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 met 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 [80]. Studies have demonstrated that the portrayal of suicide in media and online content can increase suicide rates [16] as well as how suicide-related information retrieved from SE can have either a positive or negative impact on searchers [80]. 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 suggest 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 met 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 change in the portion of MHD terms changes. Once Kristoff, who likely had his decision-making process altered due to the shifting of emotions [85], 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 not have found what he was looking for. Still, 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.*

With our example, which takes empirical findings from theory to practice, we aimed to not only convey how SE respond to users with MHD but also show the possible consequences of MHD users 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.

5 CONCLUSION, LIMITATIONS, AND FUTURE WORK

In this manuscript, we discuss the empirical analysis we conducted 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 scrutinize the subliminal stimulus that each SE under study portrays at different stages of the 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, who serve as a control group. This enables 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.

Some of the outcomes emerging from our analysis were anticipated due to the affect and language associated with the MHD under study. Still, 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 and then decline throughout the ISP. The same is not mirrored in responses to traditional users, as the stimuli SE convey to them is more balanced emotionally and contains far fewer indicators of MHD. 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-10 responses. Therefore, the very first response users see is the most stimulating and the most likely to alter users' state of mind.

To the best of our knowledge, the analysis we presented in this work is novel; 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. Presented results align with prior research highlighting 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 [39]. Moreover, our study extends that of Kazai et al. [39], who create emotional profiles of SE for traditional searchers, by including MHD indicators and emotions, as well as considering additional SE functionality and other commercial SE. Lessons learned from the presented analysis are in line with those demonstrated by Landoni et al. [42], who examined the affect of SE responses to children, regarding the need for considering the SE design and evaluating the affective dimension.

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 [55]. Our preliminary findings emphasize the value of considering searchers with MHD and their needs in the design, development, and evaluation of search tools. Given how in-vogue, and justifiably so, research pertaining to fairness in artificial intelligence is [14, 31], it is expected for more research to focus on the algorithmic side of SE for MHD searchers. Particularly as efforts thus far have primarily showcased biases of gender [12], or even clinical terms [91], in the text but little attention has been paid to MHD or text specific to SE which is worrying in light of the outcomes and implications enumerated in this work. Based on the results of our analysis, the terminology used by MHD searchers must become 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 [83], and from our results, it seems that MHD indicators 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, are limited [3]. Our analysis 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 [41]: affect does matter throughout ISP stages. Leveraging lessons learned from our study, future studies could expand theoretical ISP frameworks by integrating more stimuli cues from both a system and user perspective. Further, 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 as well as intentions in search sessions, it is clear that SE convey emotions to users consistently.

As with any research study, we identify some limitations. Although we consider the two most popular commercial SE, they are not the only ones MHD searchers turn to. Thus, future work includes conducting a similar analysis based on different languages and locations, as MHD are expressed and treated differently in other cultures. We also aim to extend the analysis to other forms of MHD, like users affected by post-traumatic stress disorder who could also be exposed to harmful stimuli through SE responses, as there is no guarantee that responses would be the same across MHD, it is important to examine beyond depression and anxiety. We also know from clinical psychology that how MHD present in children is different from adults, so considering children is another path for future exploration. One of the hurdles to expanding the current exploration is data. We created synthetic query logs to enable analysis. While standard in this type of work, the manner in which queries are generated (e.g. using word lists, query auto completion, and phrase extraction from social platforms, as we do in this work) can and does impact the data collection process. For this reason, it would be worth dedicating resources to collecting real query logs from users. This would provide more fine-grained information, like click-through and query reformulation, that can be used for interpreting how SE respond to users. Surveying users would also open the door for investigations into how users are affected by SE. Additionally, lexicons built for our analysis cannot encompass every aspect of MHD, which lead us to rely on natural language processing strategies as well. While the lexicon-based strategies we adopt resemble those common among scholarly works analyzing social media, they are not state-of-the-art among research developments in natural language processing. However, existing strategies for detecting MHD in users using persuasive technologies are generally designed and trained with social media platforms in mind, which differ heavily from how users utilize SE. Thus, future explorations of MHD detection strategies for SE, specifically ones that do not require query logs as that data is not always available, are another possible avenue for future research.

Reflecting on the outcomes of our empirical analysis, we unearthed open research directions. Is the way SE respond 'good' for users with MHD? How should SE respond to these users? Most of our profiles had significant changes between MHD and traditional users, still, do users notice the changes in the stimuli presented to them? SE are the first place most people turn to for information about almost anything. What if a user with a MHD searches for information related to their symptoms and they are met with triggering information or given access to ideas that could ultimately lead them to harm themselves? SE use 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 query 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. 2019. Search Engine Statistics. Available at: https://99firms.com/blog/search-engine-statistics/. (accessed March 20, 2020).

[2] Mental Health America. 2020. The State Of Mental Health In America. Available at: https://www.mhanational.org/issues/state-mental-health-america. (accessed August 18, 2020).

[3] Ioannis Arapakis, Joemon M Jose, and Philip D Gray. 2008. 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*. 395–402.

[4] American Psychological Association. 2020. APA Dictionary of Psychology. Available at: https://dictionary.apa.org/affect. (accessed August 21, 2020).

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

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

[7] John W Ayers, Eric C Leas, Derek C Johnson, Adam Poliak, Benjamin M Althouse, Mark Dredze, and Alicia L Nobles. 2020. Internet Searches for Acute Anxiety During the Early Stages of the COVID-19 Pandemic. *JAMA Internal Medicine* 180, 12 (2020), 1706–1707.

[8] Ion Madrazo Azpiazu, Nevena Dragovic, Oghenemaro Anuyah, and Maria Soledad Pera. 2018. 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)*. 92–101.

[9] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. 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)*, Vol. 10. 2200–2204.

[10] Amanda Baker, Naomi Simon, Aparna Keshaviah, Amy Farabaugh, Thilo Deckersbach, John J Worthington, Elizabeth Hoge, Maurizio Fava, and Mark P Pollack. 2019. Anxiety Symptoms Questionnaire (ASQ): development and validation. *General Psychiatry* 32, 6 (2019).

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

[12] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 610–623.

[13] Jared D Bernard, Jenna L Baddeley, Benjamin F Rodriguez, and Philip A Burke. 2016. 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 (2016), 317–326.

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

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

[16] 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. 2020. 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 (2020), 236–243.

[17] Andrew Campbell, Brad Ridout, Melina Linden, Brian Collyer, and John Dalgleish. 2018. 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 (2018), 208–221.

[18] Stevie Chancellor and Munmun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine* 3, 1 (2020), 1–11.

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

[20] Paula Glenda Ferrer Cheng, Roann Munoz Ramos, Jó Ágila Bitsch, Stephan Michael Jonas, Tim Ix, Portia Lynn Quetulio See, and Klaus Wehrle. 2016. Psychologist in a pocket: lexicon development and content validation of a mobile-based app for depression screening. *JMIR mHealth and uHealth* 4, 3 (2016), e88.
[21] J. Clement. 2019. 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/. (accessed March 20, 2020).
[22] Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. 2015. CLPsych 2015 shared task: Depression and PTSD on Twitter. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*. 31–39.
[23] Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth international AAAI conference on weblogs and social media*.
[24] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*.
[25] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. 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*. 2098–2110. https://doi.org/10.1145/2858036.2858207
[26] Gianluca Demartini and Stefan Siersdorfer. 2010. 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*. 7. https://doi.org/10.1145/1863879.1863883
[27] Jan B Engelmann, Todd A Hare, Andrew S Fox, Regina C Lapate, Alexander J Shackman, and Richard J Davidson. 2018. Emotions can bias decision-making processes by promoting specific behavioral tendencies. (2018).
[28] BJ Fogg. 2003. *Persuasive technology: using computers to change what we think and do*. Morgan Kaufmann.
[29] Tatiana Gossen. 2016. *Search engines for children: search user interfaces and information-seeking behaviour*. Springer.
[30] Jamie Guillory, Jason Spiegel, Molly Drislane, Benjamin Weiss, Walter Donner, and Jeffrey Hancock. 2011. Upset now? Emotion contagion in distributed groups. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 745–748.
[31] Anhong Guo, Ece Kamar, Jennifer Wortman Vaughan, Hanna Wallach, and Meredith Ringel Morris. 2019. Toward fairness in AI for people with disabilities: A research roadmap. *arXiv preprint arXiv:1907.02227* (2019).
[32] Max Hamilton. 1959. The assessment of anxiety states by rating. *British Journal of Medical Psychology* 32, 1 (1959), 50–55.
[33] Max Hamilton. 1967. Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology* 6, 4 (1967), 278–296.
[34] Markham Heid. 2018. You Asked: Is It Bad for You to Read the News Constantly? Available at: https://time.com/5125894/is-reading-news-bad-for-you/. (accessed March 18, 2020).
[35] Michael Herdman, Claire Gudex, Andrew Lloyd, MF Janssen, Paul Kind, David Parkin, Gouke Bonsel, and Xavier Badia. 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research* 20, 10 (2011), 1727–1736.
[36] Nicholas C Jacobson, Damien Lekkas, George Price, Michael V Heinz, Minkeun Song, A James O'Malley, and Paul J Barr. 2020. 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 (2020), e19347.
[37] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2017. Accurately interpreting clickthrough data as implicit feedback. In *ACM SIGIR Forum*, Vol. 51. Acm New York, NY, USA, 4–11.
[38] Koenigbauer Josephine, Letsch Josefine, Doebler Philipp, Ebert David, and Baumeister Harald. 2017. Internet-and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. *Journal of affective disorders* 223 (2017), 28–40.
[39] Gabriella Kazai, Paul Thomas, and Nick Craswell. 2019. The Emotion Profile of Web Search. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1097–1100.
[40] Adam DI Kramer. 2012. The spread of emotion via Facebook. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 767–770.
[41] Carol C Kuhlthau. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science* 42, 5 (1991), 361–371.
[42] Monica Landoni, Maria Soledad Pera, Emiliana Murgia, and Theo Huibers. 2020. 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*. 136–144. https://doi.org/10.1145/3340631.3394847
[43] Eleanor Leigh and David M Clark. 2022. Establishing the psychometric properties of the self-report Liebowitz Social Anxiety Scale for Children and Adolescents in a general population adolescent sample. *European Journal of Psychological Assessment* 38, 4 (2022), 243.
[44] Michael R Liebowitz. 1987. Social phobia. *Modern problems of pharmacopsychiatry* (1987).

[45] Chenhao Lin, Pengwei Hu, Hui Su, Shaochun Li, Jing Mei, Jie Zhou, and Henry Leung. 2020. Sensemood: Depression detection on social media. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*. 407–411.

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

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

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

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

[50] Daniel M Low, Laurie Rumker, Tanya Talkar, John Torous, Guillermo Cecchi, and Satrajit S Ghosh. 2020. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. *Journal of medical Internet research* 22, 10 (2020), e22635.

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

[52] Myfanwy Maple, Julie Cerel, Rebecca Sanford, Tania Pearce, and Jack Jordan. 2017. Is exposure to suicide beyond kin associated with risk for suicidal behavior? A systematic review of the evidence. *Suicide and Life-Threatening Behavior* 47, 4 (2017), 461–474.

[53] Patricia Martın-Rodilla, David E Losada, and Fabio Crestani. 2022. Overview of eRisk 2022: Early Risk Prediction on the Internet. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 13th International Conference of the CLEF Association, CLEF 2022, Bologna, Italy, September 5–8, 2022, Proceedings*, Vol. 13390. Springer Nature, 233.

[54] Thomas F Martinelli, Gert-Jan Meerkerk, Gera E Nagelhout, Evelien PM Brouwers, Jaap van Weeghel, Gerdien Rabbers, and Dike van de Mheen. 2020. 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 (2020), 1504–1513.

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

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

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

[58] Yashar Moshfeghi and Joemon M. Jose. 2013. On Cognition, Emotion, and Interaction Aspects of Search Tasks with Different Search Intentions. In *Proceedings of the 22nd International Conference on World Wide Web*. 931–942. https://doi.org/10.1145/2488388.2488469

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

[60] Fedelucio Narducci, Marco De Gemmis, and Pasquale Lops. 2015. 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*. 3–6.

[61] Usman Naseem, Adam G Dunn, Jinman Kim, and Matloob Khushi. 2022. Early identification of depression severity levels on reddit using ordinal classification. In *Proceedings of the ACM Web Conference 2022*. 2563–2572.

[62] Michelle G Newman, Andrea R Zuellig, Kevin E Kachin, Michael J Constantino, Amy Przeworski, Thane Erickson, and Laurie Cashman-McGrath. 2002. 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 (2002), 215–233.

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

[64] Javier Parapar, Patricia Martín-Rodilla, David E Losada, and Fabio Crestani. 2021. Overview of erisk 2021: early risk prediction on the internet. In *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, 324–344.

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

[66] Rebecca M Puhl. 2020. 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 (2020), e13008.
[67] Chengcheng Qu, Corina Sas, Claudia Daudén Roquet, and Gavin Doherty. 2020. Functionality of top-rated mobile apps for depression: Systematic search and evaluation. *JMIR mental health* 7, 1 (2020), e15321.
[68] Esteban A Ríssola, Mohammad Aliannejadi, and Fabio Crestani. 2022. Mental disorders on online social media through the lens of language and behaviour: Analysis and visualisation. *Information Processing & Management* 59, 3 (2022), 102890.
[69] Esteban A Ríssola, David E Losada, and Fabio Crestani. 2019. Discovering Latent Depression Patterns in Online Social Media. In *Proceedings of the 10th Italian Information Retrieval Workshop* (Padova, Italy) *(IIR '19)*. 13–16. http://ceur-ws.org/Vol-2441/paper10.pdf
[70] Esteban A. Ríssola, David E. Losada, and Fabio Crestani. 2021. A Survey of Computational Methods for Online Mental State Assessment on Social Media. *ACM Trans. Comput. Healthcare* 2, 2, Article 17 (March 2021), 31 pages. https://doi.org/10.1145/3437259
[71] Farig Sadeque, Dongfang Xu, and Steven Bethard. 2018. Measuring the latency of depression detection in social media. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 495–503.
[72] Saif Saif Mohammad. 2018. Word Affect Intensities. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*.
[73] H Andrew Schwartz, Johannes Eichstaedt, Margaret Kern, Gregory Park, Maarten Sap, David Stillwell, Michal Kosinski, and Lyle Ungar. 2014. 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*. 118–125.
[74] Sima Sharifirad, Borna Jafarpour, and Stan Matwin. 2019. 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).
[75] Dushyant Sharma, Rishabh Shukla, Anil Kumar Giri, and Sumit Kumar. 2019. A Brief Review on Search Engine Optimization. In *Proceedings of the 9th International Conference on Cloud Computing, Data Science and Engineering (Confluence)*. IEEE, 687–692.
[76] Judy Hanwen Shen and Frank Rudzicz. 2017. Detecting anxiety through reddit. In *Proceedings of the 4th Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*. 58–65.
[77] Anu Shrestha and Francesca Spezzano. 2019. Detecting Depressed Users in Online Forums. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 945–951. https://doi.org/10.1145/3341161.3343511
[78] Robert L Spitzer, Kurt Kroenke, Janet BW Williams, Patient Health Questionnaire Primary Care Study Group, Patient Health Questionnaire Primary Care Study Group, et al. 1999. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Jama* 282, 18 (1999), 1737–1744.
[79] Helen M Startup and Thane M Erickson. 2006. The penn state worry questionnaire (PSWQ). *Worry and its psychological disorders: Theory, assessment and treatment* (2006), 101–119.
[80] Benedikt Till, Marlies Braun, Susanne Gahbauer, Nikolaus Reisinger, Ewald Schwenzner, and Thomas Niederkrotenthaler. 2020. 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 (2020), 300–309.
[81] William Tov, Kok Leong Ng, Han Lin, and Lin Qiu. 2013. Detecting well-being via computerized content analysis of brief diary entries. *Psychological Assessment* 25, 4 (2013), 1069.
[82] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. Recognizing Depression from Twitter Activity. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3187–3196. https://doi.org/10.1145/2702123.2702280
[83] Frans van der Sluis and Betsy van Dijk. 2010. 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)*.
[84] Nikhita Vedula and Srinivasan Parthasarathy. 2017. Emotional and Linguistic Cues of Depression from Social Media. In *Proceedings of the 2017 International Conference on Digital Health*. 127–136. https://doi.org/10.1145/3079452.3079465
[85] Kirsten G Volz and Ralph Hertwig. 2016. Emotions and decisions: Beyond conceptual vagueness and the rationality muddle. *Perspectives on Psychological Science* 11, 1 (2016), 101–116.
[86] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. 2018. 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, Article 43 (March 2018), 26 pages. https://doi.org/10.1145/3191775
[87] Luyan Xu, Xuan Zhou, and Ujwal Gadiraju. 2019. 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*. 1249–1252. https://doi.org/10.1145/3331184.3331353

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

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

[90] Anis Zaman, Rupam Acharyya, Henry Kautz, and Vincent Silenzio. 2019. Detecting Low Self-Esteem in Youths from Web Search Data. In *The World Wide Web Conference*. 2270–2280. https://doi.org/10.1145/3308558.3313557

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

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

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

[94] Ayah Zirikly, Philip Resnik, Ozlem Uzuner, and Kristy Hollingshead. 2019. CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology*. 24–33.