Not Just Skipping: Understanding the Effect of Sponsored Content on Users’ Decision-Making in Online Health Search

Anat Hashavit
Bar Ilan University, Israel
anat.hashavit@gmail.com

Hongning Wang
University of Virginia, Charlottesville, VA, USA
hw5x@virginia.edu

Tamar Stern
Bar Ilan University, Israel
stern.tamar96@gmail.com

Sarit Kraus
Bar Ilan University, Israel
sarit@cs.biu.ac.il

ABSTRACT
Advertisements (ads) are an innate part of search engine business models. To promote sales, advertisers are willing to pay search engines to promote their content to a prominent position in the search result page (SERP). This raises concerns about the search engine manipulation effect (SEME); the opinions of users can be influenced by the way search results are presented.

In this work, we investigate the connection between SEME and sponsored content in the health domain. We conduct a series of user studies in which participants need to evaluate the effectiveness of different non-prescription natural remedies for various medical conditions. We present participants SERPs with different intentionally created biases towards certain viewpoints, with or without sponsored content, and ask them to evaluate the effectiveness of the treatment solely based on the information presented to them. We investigate two types of sponsored content: 1) Direct marketing ads that directly market the product without expressing an opinion about its effectiveness; and 2) Indirect marketing ads that explicitly advocate the product’s effectiveness on the condition in the query. Our results reveal a significant difference between the influence on users from these two types of ads. Though users mostly skip direct marketing ads, they do sometimes tilt users’ decision-making. Indirect marketing ads affect both the users’ examination behavior and their perception of the treatment’s effectiveness. We further discover that the contrast between the indirect marketing ads and the viewpoint presented in the organic search results plays an important role in users’ decision-making. When the contrast is high, users exhibit a strong preference towards a negative viewpoint, and when the contrast is low or none, users exhibit a preference toward a more positive viewpoint.

CCS CONCEPTS
• Human-centered computing → User studies; • Information systems → Information retrieval; • Applied computing → Health informatics.

KEYWORDS
user study, information retrieval, bias

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference Format:
Anat Hashavit, Hongning Wang, Tamar Stern, and Sarit Kraus. 2023. Not Just Skipping: Understanding the Effect of Sponsored Content on Users’ Decision-Making in Online Health Search. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ’23), July 23–27, 2023, Taipei, Taiwan. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3539618.3591744

1 INTRODUCTION
Search engines are a gateway to information, which help users make informed decisions in various aspects of their lives. This includes decisions about their health [2]. Users trust that search engines will present them with accurate and reliable results [1]. However, this trust can make them susceptible to biases in the search results: whether it is the general and often positive bias of the presented results [45, 47]; or the users’ own tendency to inspect mostly the top results on the page [25, 27, 35, 36]. In the health domain in particular, it has been shown that users are mainly susceptible to positive bias when researching medically related topics [5, 19, 39, 45]. This problem can be exacerbated when financial and marketing influences come into play, if, for example, people count on natural remedies or vitamins as an alternative to professional medical treatment or vaccinations. This is a concerning phenomenon especially now during the COVID pandemic, when consuming inaccurate medical information can be life threatening.

On the other hand, search engines are also a powerful advertising platform where the advertisers can connect their content to relevant queries, thus specifically targeting relevant audiences [43]. Search engine advertising has become very common and users are accustomed to encounter ads in their SERP. As a result, they sometimes develop a tendency to disregard the ads: skipping over them as if they do not exist, a phenomenon often referred to as banner blindness [7, 14, 31]. Nevertheless, marketing research shows search engine advertising is an effective tool advancing various consumer metrics, such as brand awareness and brand consumption, even among users who do not click on ads [51].

These two sets of seemingly contradictory results motivate our research about the connection between SEME and sponsored content in healthcare domain. As sponsored content can introduce additional bias to the presented results, it can present a potential issue when the search process is part of a user’s decision-making process in health-related problems. Take for example a parent who has a child with attention deficit hyperactive disorder (ADHD) and wishes to explore options of natural-based remedies for treatment of ADHD. The parent starts their research with their trusted search engine. Figure 1 shows the top of the result page for the query “natural remedies for ADHD”. As can be seen, the first viewpoint of the SERP is almost completely dominated by sponsored content. The
The vast majority of the sponsored content is highly relevant to ADHD and promotes natural non-prescription products, whose names strongly suggest benefits to attention and focus. Unfortunately, the fact is that most of these products have not been scientifically proven to be effective [16, 20], but this information is not presented next to the ad. Although this information might be included in one of the returned results lower on the page, the parent might not notice them as extra effort is needed to examine those positions [25, 50]. This bias can prevent the parent from making an informed decision that considers all relevant information, which in turn can harm their child [18].¹

Previous work showed that users tend to prefer organic content to sponsored content [11, 23, 32]. In this study, we investigate whether the same conclusion can be reached in our domain of interest, and explore the effect of sponsored content might have on users’ decision making in this domain; in particular, under the task setting of determining the effectiveness of online available natural remedies. We focus our study on the following research questions:

- **RQ1**: When users interact with a search engine to decide the effectiveness of a medical treatment, how is users’ result examination behaviour affected by the sponsored content on a SERP?
- **RQ2**: When users interact with a search engine to decide about the effectiveness of a medical treatment, does the existence of sponsored content introduce bias to their decisions?

In order to answer these research questions, we conducted a series of user studies in which participants were presented with SERPs discussing the effectiveness of natural remedies on corresponding medical conditions. Participants were presented with a SERP that only included results that specifically discuss the related query but with different viewpoints (e.g., for vs. against). The participants were then requested to only use the results presented to them to decide the effectiveness of the remedy mentioned in the query. Each SERP contained an equal number of organic results from each viewpoint, but the order in which the results were presented varied so that participants were exposed to different levels of rank bias. In the meanwhile, the participants were presented with a SERP that 1) did not contain any ads, or 2) with an ad that directly markets the remedy discussed in the query, or 3) with an indirect marketing ad, which, instead of directly offering a product for purchase, refers to the promotional content discussing the benefit of the product for the purpose of building the brand. We tracked the participants’ click behaviour and their responses to our post-study survey questions and compared the results under various settings.

We found that participants tended to skip direct marketing ads and proceed to organic content. However, the existence of the ads could still affect users’ exploration behaviour and as a result their decision making. For indirect marketing ads, we found that the effect of ads on both participants’ examination behaviour and their responses depended on the ranking bias of the organic results. When the organic results were biased towards a negative viewpoint, participants not only skipped the indirect marketing ads during their search but also enforced their negative bias towards the treatment’s effectiveness. When the organic results were biased towards an inconclusive or positive viewpoint, however, participants were more inclined to inspect the content of the ads which introduced positive bias to their decision making.

## 2 BACKGROUND AND HYPOTHESIS

We explore the connection between SEME and sponsored content in the health search domain. In this section, we first discuss existing work related to all these aspects and then present the hypotheses related to our research questions.

**Impact on Users’ Beliefs.** Most studies that explored search engines’ effect on users’ beliefs and decision-making show that higher ranked results are more likely to affect the users, especially when the results are biased towards a positive viewpoint [5, 12, 39, 46].

In the healthcare domain, White and Hassan [47] showed that the search results in medical related queries are often biased toward positive outcomes. White [45] investigated the bias issue in yes-no questions based on search results about medical conditions. Similar results are observed, showing that users favored more positive results irrespective of the truth. More so, White showed that almost half of the times the viewpoints presented in the search results were scientifically incorrect. Bink et al. [6] conducted a study on the influence of featured snippets on users’ assessments of the effectiveness of medical treatments for various conditions. The results demonstrated that users may be vulnerable to misinformation in these types of snippets. Pogacar et al. [39] presented users with biased SERPs and showed that this bias affected users’ decision-making in health related questions, and Hashavít et al. [19] showed that this phenomena repeats even when the users are free to conduct the search on their own.

**Impact on Users’ Search Experience and Behaviour.** Advertisements’ effect on users’ search experience has been researched in several studies. Lewandowsk et al. [29, 30] conducted two large-scale studies on German internet users that focused on Google. Their studies found that users had limited knowledge regarding a search engine’s business model. Many of them reported that they were either unaware that it was possible to pay Google to present content on the SERP or they did not know how to distinguish between ads and organic results. Schultheiß and Lewandowski [41] showed that users with less understanding of search engine business models were more likely to click on ads. Several studies found

¹The effective course of treatment for ADHD is out of the scope of this paper.
that the quality of ads and their rank affected the attention that ads received as well as the probability the user would click on them [4, 8, 11, 14, 17, 28, 38]. This correlates with many studies showing users exhibit strong position bias towards higher ranked results [25, 27, 35, 36]. However, when examining attention to ads compared to organic results, studies found negative bias toward the sponsored content. Ads received significantly less attention from users in comparison to organic results, and on several occasions they distracted and frustrated the users [8, 11, 14, 17] and overall had a negative impact on users’ search experience and their performance in the search task.

Much research attention has been paid to sponsored content’s effectiveness under various conditions from a marketing perspective. The goal of these studies is mainly to examine factors that make advertisement more effective. The main metric inspected is click through rate [15, 24, 32], some inspect users’ attitude towards online advertising [33] and percentage of successful purchases originated from ads [3]. These studies focus mainly on e-commerce and do not explore the effect of sponsored content on users’ beliefs. These studies as well find consistent preference of organic results in comparison to sponsored content, even for high ranked results [23]. Studies also found an interdependence between organic and sponsored results: the existence of sponsored content increases the probability of an organic result from the same domain to be inspected [40, 48]. Indirect marketing advertising, also referred to as native advertising, is also a topic of interest to the marketing research community. These types of ads are designed to look and feel similarly to the organic content surrounding them in an attempt to minimize their disruption to the user experience, whether in a search engine, a news website, or a social network news feed. While there is a debate between researchers as to the effectiveness of this tool, there is an agreement that consumers are able to distinguish between native ads and organic content [40, 44, 49].

To summarize, existing studies examined the effect of ads on users’ behaviour and search experience from multiple angles but they did not touch the effect of sponsored content on users’ perception or their beliefs, especially not in the health domain. Previous work in other domains suggests users are prone to prefer organic to sponsor content. Our research investigates if this pattern repeats in the health domain. We explore the effect that ads might have on users’ decision-making; or more specifically, their beliefs and biases, and their perception of truth when their information need relates to the effectiveness of medical treatments.

We therefore phrase two null hypotheses in accordance to our research questions in the context of health-related search:

- **H1:** when users interact with a search engine to decide about the effectiveness of a medical treatment and sponsored content is presented in the SERP, users will generally skip it and examine mostly organic results.
- **H2:** when users interact with a search engine to decide the effectiveness of a medical treatment, the existence of sponsored content will not introduce any bias to their decisions.

## 3 METHOD AND EXPERIMENTAL SETUP

We designed a series of user studies in which participants were required to evaluate the effectiveness of an intentionally selected natural remedy in treating a particular medical condition, only based on the results in a SERP presented to them.

To this end, we created a dataset of SERPs with a wide variety of viewpoint biases (e.g., supporting or against the natural remedy’s effectiveness), different configurations of the SERP (e.g., ranking positions of organic search results, with/without ads) and queries related to different remedies and health conditions. Data regarding users’ examination behaviour was collected by tracking users’ clicks using JavaScript during the study. Insight regarding users’ decision making and their search experience was collected explicitly via a series of survey questions by the end of the study.

### 3.1 Dataset Preparation

#### 3.1.1 Queries

We collected a set of SERPs related to various queries about natural remedies’ effectiveness for a certain health condition. We chose the following three queries: **Q1:** Ginkgo biloba for treating tinnitus; **Q2:** Melatonin for treating jetlag; **Q3:** Omega fatty acids for treating ADHD.

Each of these three queries has a different ground-truth answer. According to their corresponding Cochrane reviews: Ginkgo biloba is **not effective** in treating tinnitus [22]; Melatonin **is effective** in treating jetlag [21]; and for Omega-3 the answer is inconclusive and further study is required in order to reach a conclusion [16]. The detailed search result entries were manually retrieved by searching the above queries in well-known search engines (i.e., Google and Bing). The query phrases entered to the search engines followed patterns that were designed to retrieve relevant results that contain a viewpoint regarding the effectiveness of the treatment for the condition discussed in the query. For example, “is melatonin effective for jetlag”,” is melatonin ineffective for jetlag”, etc.

#### 3.1.2 Advertisements

A SERP can have one of three possible ad configurations: 1) Ads clean - no sponsored content at all; 2) Direct marketing ads; and 3) Indirect marketing ads.

Direct marketing ads are ads that offer the remedy in question for sale. These ads refer the users to an online shopping website where they can purchase the remedy. The shopping websites often do not directly discuss the effectiveness of the remedy, in particular not in the title or the snippet of the ads. Several examples of such ads can be seen in Figure 1. We hypothesize that the existence of such results might indirectly suggest to users that the product is effective. Direct marketing ads were retrieved from actual SERPs returned by the commercial search engines.

Indirect marketing ads are entries of sponsored content that broadly discusses and supports the use of the remedy for the condition in question. Indirect marketing ads highlight the treatment’s effectiveness already in their titles and snippets. For example, an indirect marketing ad for omega fatty acids had the following title: “Omega 3s: The Ultimate (ADHD) Brain Food - ADDitude”, and the following snippet: “Supplementing with omega-3s eases hyperactivity; Analyzing data from 16 studies on ADHD and omega-3s, researchers at Oregon Health & Science ...” Although these web pages do not focus on selling the product directly, all of them have links on their websites referring users to a shopping website. Indirect marketing ads are also a notable occurrence in SERPs returned by commercial search engines.
ads were collected from organic search results that were annotated as having a positive viewpoint and which contained links to sales pages of the treatment in question as part of their web page. Notice that we excluded third party ad space such as Google AdSense but only considered static commercial content for the treatment in question when classifying indirect marketing ads.

Our SERPs with sponsored content only included one ad (either direct or indirect marketing ad) placed at the top of the ranked list. Ads were marked by the bold text Ad to the left of their titles, as most commercial search engines would do. The manufactured SERPs resided on our experiment servers but the links to which the users were referred point to actual web pages found online, which we did not alter. Since our study’s research questions focus on users’ perception but not their general search experience, and given the high relevance of commercial search engines in presenting related ads to users, we opted to present only ads that were relevant to the search query at hand.

### 3.1.3 Viewpoint Bias.

The viewpoint of each entry was independently evaluated by two readers with necessary medical background knowledge. Entries on which the readers disagreed were removed from the dataset. We categorize possible viewpoints into three classes: supporting (Y), rejecting (N) and inconclusive (M), in order to coincide with the options later presented to our study participants.

All SERPs had the same number of entries for each of three possible viewpoints to avoid introducing other possible biases. However, they differed in the level of ranking bias induced by their actual rank positions. We intentionally created two forms of ranking bias: block-wise ranking bias and interleaved ranking bias. In block-wise ranking bias, the viewpoints were divided into three blocks of consecutive result rankings about the same viewpoints. For example, the configuration \(yyymmmmnnn\) specifies a SERP with 9 results, where three supporting entries are placed in rank position 1-3, three inconclusive results in rank positions 4-6, and three rejecting entries in rank positions 7-9. In interleaved ranking bias, the viewpoints were interleaved with each other. For example, in the configuration \(ymmymnymn\) three supporting entries are placed in rank positions \(1,4,7\), three inconclusive results in rank positions \(2,5,8\) and three rejecting entries in rank positions \(3,6,9\). We denote a sequence of \(a\) consecutive viewpoints of value \(v\) as \(v^a\), and a sequence of interleaved viewpoints \(a_1\ldots a_\ell\) of length \(a\) as \((a_1\ldots a_\ell)^a\).

For example, the sequence \(yyymmmmnnn\) will be denoted as \(y^3m^3n^3\) and the sequence \(ymnymnym\) will be denoted as \((ynyn)^3\).

The total a-priori ranking bias of a viewpoint in a SERP can be viewed as the expected attention that the viewpoint receives. Formally, the ranking bias that viewpoint \(v\) expects to receive in a given result ranking \(D\) of length \(R\) is defined as:

\[
B(D, v) = \frac{1}{Z} \sum_{i=1}^R P_e(i) \mathbb{1}_{D[i]=v}
\]

where \(Z\) is a normalization constant, \(\mathbb{1}_{D[i]=v}\) suggests if the document placed at rank \(i\) corresponds to viewpoint \(v\) and \(P_e(i)\) is the probability of the document to be examined by a user. Given no prior knowledge, we follow the position-based examination model [10] to estimate the probability by \(P_e(i) = \frac{1}{R}\).

Table 1 describes the possible SERP configurations grouped according to the viewpoint which receives the maximal ranking bias. The first four, the middle four, and the last four are sequences where the maximal bias belongs to the supporting, inconclusive, and rejecting viewpoints respectively. The values in columns B(Y), B(M) and B(N) show the ranges of prior and posterior (in parenthesis) bias levels for the configuration associated with each maximal bias viewpoint, as was computed according to Eq (1). The posterior bias values will be discussed in the experiment results section; they were calculated based on the actual click-through rate observed in our user studies.

In our study the value \(a\) was set to either 2 or 3 depending on the number of retrieved result entries for each viewpoint. For query Q1 (‘Ginkgo-biloba for treating tinnitus’), we could not find three entries that expressed an inconclusive viewpoint, since its ground-truth is not effective; therefore in order to balance viewpoints, all SERPs for Q1 were limited to 6 results, i.e., \(a = 2\). For the other queries, \(a\) was set to 3. For Q2 (‘Melatonin for treating jetlag’), we could not find an indirect marketing ad entry. We assume this is because Melatonin can be easily purchased over the counter (at least in the US). Therefore our indirect marketing content experiments were only conducted for query Q1 and Q3.

For each query there were 12 different bias configurations. For each query-configuration pair, three SERPs were randomly generated. To create the SERP instances with ads, a relevant ad of each query-configuration pair, three SERPs were randomly generated. These basic statistics about our study dataset are summarized in Table 2. 3

### 3.2 Variables.

The dependant variable in this study is the participants’ perceived effectiveness of a natural remedy for a given health condition. It is a categorical variable with three levels: effective, ineffective and inconclusive. The independent variables manipulated in our experiments were as follows:

1. (1) The discussed query (categorical; between-subjects). Each SERP is about a single query discussing the effectiveness of one treatment for a specific condition.

2. (2) The sponsored content presented in the SERP (categorical; between-subjects). A SERP could contain a direct marketing ad, indirect marketing ad, or no ads at all.

\[\text{The data set will be published and available to future research.}\]
In this section, we describe the procedure of our user study. We used the Amazon Mechanical Turk (MTurk) platform to recruit participants for our study. All participants were from English speaking countries. We used MTurk’s external question format which directed participants to the website hosted by our survey web server. Prior to the beginning of our study, we obtained an institutional review boards (IRB) approval from our institute.

Upon entering the survey, each participant was requested to enter demographic details, which include age, gender, education level and field of education. After providing their personal information, the participant was directed to the instruction page. At this point, participants were introduced to the query assigned to them (randomly) and their task was explained.

After reading the instructions, participants were asked if they had any previous knowledge regarding the topic of the study. We did not want to collect responses that were resulted from the participants’ prior knowledge or bias, but we also did not want them to lie in order to participate for payment. Hence, we allowed all participants to take the rest of the survey, but excluded responses from those who declared having prior knowledge about their assigned query from our analysis. Whether their responses were included in our analysis did not affect how they got paid in our study.

After answering this question, the participants were referred to the assigned SERP to start the task. They were instructed to browse as many links as they needed to answer the question; and they could not progress to the next stage unless they had clicked at least one link. This was to ensure the participants understood that they had to interact with the SERP.

Once the participants pressed the “Answer Survey” button that was located next to the search box, they were directed to the question form, where they were required to enter their conclusion regarding the effectiveness of the treatment discussed in the query. The possible answers were: Yes, No, Maybe and Not Sure. In this form, we also included 3 attention questions to filter out careless workers. Participants were required to 1) answer a multiple choice question about the nature of the remedy that the query was intended for, 2) answer a multiple choice question about the nature of the medical condition that the query concerned, and 3) provide a reason or piece of evidence for their conclusion regarding the remedies’ effectiveness.

The attention question about the nature of the remedy was phrased simply as ‘What is REMEDY’. The available answers were: 1) A hormone (true for melatonin); 2) A nutrient found mostly in fish, nuts and seeds (true for omega fatty acids); and 3) A tree (true for Ginkgo biloba). Similarly, the attention question about the nature of the medical condition was phrased as ‘What is MEDICAL CONDITION’. The available answers were: 1) A neurological disorder (true for ADHD); 2) Ringing noise in one or both ears (true for tinnitus); and 3) A sleep problem (true for jet-lag).

Participants who did not answer the filtering questions correctly or did not provide a coherent reason for their choice were removed from the analysis. Participants were also provided with the opportunity to comment about the study itself. By the end of our study, we received very positive feedback from the participants, indicating the survey was clear to follow and the task was easy to perform.

Once participants submitted their answers about the treatments’ effectiveness and attention questions, they could not go back and alter them. If the SERP presented to the participants included ads, they were presented with two additional questions about the influence of the ad to their overall search experience and to their decision about the treatment’s effectiveness. Participants were also allowed to enter any general comments concerning the ad. The first question was phrased as: “The search result page you viewed contained sponsored content (Ads). How much did the ad affect your search experience?” The responses to this question were measured by a 5-point Likert scale which ranged from -2 to +2, with negative values indicating the ads worsened the search experience and positive values indicating it improved it. The second question was phrased as: “How much did the ad affect your decision regarding the treatment’s effectiveness?” We again used a 5-point Likert scale with negative values indicating the ads added doubts to participants’ decisions and positive values indicating ads enhanced their belief in their decision.

Once the participants completed the task, they would receive a verification code to receive their payment from the MTurk website. On this page, they were also presented with a highlighted warning message noting that they were participating in an experiment and should not make any conclusions about the effectiveness of the remedy discussed in the query. Since our study is between subject, each participant was exposed to only one SERP. Once a participant completed the task, they were not offered to take the survey again.

Participants were paid at a basis rate of $1 for participation. In addition, they were informed that they would receive a $1 bonus if their answers were qualified based on their logged search behaviors. By our design, to receive the bonus they had to answer our post survey attention questions correctly and spend at least two minutes on the task itself. The payment scheme was derived from a preliminary experiment run in our lab by volunteers that were not MTurk workers, in order to estimate the time it takes to complete the survey task. We added additional time to account for slower working participants, in order to make sure we comply with an hourly payment of at least the US minimum wage. A later analysis of the average time it took the actual participants to complete the task revealed that this was indeed the case.

### 4 EXPERIMENT RESULTS

We filtered out responses from participants who answered our attention question incorrectly (92 responses). We also removed responses from users who did not provide any reason supporting

| Table 2: Statistics of the user study dataset. |
|-----------------------------------------------|
|                                | Q1 | Q2 | Q3 |
| ---                           | --- | --- | --- |
| a                             | 2   | 3   | 3   |
| # of entries in a SERP without an ad | 6   | 9   | 9   |
| # of entries in a SERP with an ad | 7   | 10  | 10  |
| # of SERP bias configurations | 12  | 12  | 12  |
| # of SERP instance per configurations | 3  | 3   | 3   |
| Total # of unique SERPs       | 108 | 72  | 108 |
Table 3: Levels and descriptive statistics of observations per independent variable.

| Independent Variable | Levels                  | #Responses |
|----------------------|-------------------------|------------|
| Ad Configuration     | No ads                  | 157        |
|                      | Direct Marketing        | 151        |
|                      | Indirect Marketing      | 155        |
| Query                | Ginkgo-Biloba for tinnitus | 172       |
|                      | Melatonin for jetlag    | 105        |
|                      | Omega fatty Acids for ADHD | 186      |
| SERP Bias            | Y                       | 152        |
|                      | M                       | 159        |
|                      | N                       | 152        |

their responses (63 responses), which is explicitly required in our instructions. Examples of bad reasons provided by users include an empty phrase or single-word sentences and vague answers such as: “it provides correct information”. Phrases only describing the treatments or the condition without referring to the effectiveness of the treatment were also removed.

After filtering, our user study produced 463 valid responses from 463 different participants. Each participant was assigned to a single SERP. The participants’ aged ranged from 20 to 74 with an average age of 39 (sd=11) and a balanced gender distribution (58% male, 41% female, 1% other). Table 3 summarizes the number of responses received for the different settings of each independent variable. Participants spent an average of 4.7 minutes on a given SERP (std=3.7) and clicked 2.3 links (std=1.35) in our study. In total, 1040 links were entered by all users, of which 972 (93.5%) were organic results and 68 (6.5%) were links to sponsored content.

4.1 Experiment Setup Validation

Before discussing our research questions, we first validate our experimental setup. The experiments were designed with the expectation that participants would reach a decision based on the information they consumed from the SERP. A valid setup should thus result in a consistent correlation between participants’ behaviour and their decisions, which consequently should allow us to use the behaviour data to predict users’ decisions. We therefore trained a multi-class logistic regression classifier and analyzed its learnt feature weights.

The classifier was fitted to predict a user’s response given the SERP presented to her based on her observed clicks. Binary features in the form of \( f_j^i = \{0, 1\} \) were constructed to encode users’ clicks, where \( i \) denotes the position of a document and \( l \) denotes the viewpoint label (“effective”, “inconclusive” or “ineffective”) or the ad configuration (“direct” or “indirect”) of that document. \( f_j^i = 1 \) if the document at position \( i \) with label \( l \) was clicked, and 0 otherwise. For example \( f_1^1 \) means that in one experiment the user clicked on the direct marketing ad at position 1.

We used \( \ell_1 \) regularization for feature selection purposes, to help the classifiers learn the most significant features. The trained classifier produced a ROC-AUC score of 0.85. In addition we ran a generalized Hosmer and Lemeshow goodness of fit test. The Hosmer and Lemeshow test compares the event rates of the models’ predicted classes to that of the ground truth. If these rates are significantly different the model is not fitted well [13]. The test produced a non significant p-value (p=0.85) indicating a good fit of the model. Table 4 reports the feature weights and weight ranges for each answer class by row (i.e., “Yes”, “Maybe”,”No”). The higher the weight value, the stronger the contribution of that feature to the classification of the corresponding class. A weight of zero indicates no contribution, and negative weights indicate a negative correlation. For example, the learnt weights for features of the form \( f_{ineffective}^i \) in the model predicting a “Yes” response range from -0.8 to -0.39, across the ranking positions. That means that clicking on an “ineffective” result acts as a counter indication to selecting “Y”.

Examining the table we can report that participants behaved according to our expectations. For each answer class only features corresponding to the examined entries with the same viewpoint received positive weights. The rest received either a negative weight or zero. We also observe that clicking on direct marketing ads has almost no effect on participants’ decision making, while inspecting indirect marketing ads has a strong correlation with participants’ decision to select the “Yes” response option, and a negative correlation to their decision to select the “No” response option.

4.2 Examination Behaviour

Figure 2 presents CTR over rank positions, for each of the three possible ad configurations. For convenience, we refer to each configuration by either one or two letters corresponding to an ad configuration and a posterior bias level. Configurations noted by a single letter (‘Y’, ‘M’, ‘N’) correspond to bias levels of ad-free SERPs, and configurations marked by two letters correspond to different bias levels for SERPs with either direct (D) or indirect (I) marketing ads. Examining the results we can immediately observe a strong preference towards top ranked organic results in all ad configurations and all bias levels, concurring with previous work on position bias. We can additionally observe that there are different behavioural patterns between ad configurations and often between different bias levels in the same ad configuration as well.

Our null hypothesis states that users will prefer organic content over sponsored content. To test this hypothesis we conducted a set of t-tests in which we compared the CTR of the ad entry to that of the first organic result. In addition, to gain some more insight, we compared the CTR of the top two organic results (i.e., ranks 1 and 2 for the no-ads configuration, and ranks 2 and 3 for the two ads configurations). We conducted these tests for all ad-bias configurations, including the no-ads configuration for reference. The results are presented in Table 5.

Examining the table we observe that while in SERPs with direct advertising users do seem to skip the ad for all SERP bias levels, for indirect marketing ads users’ attitude towards the ad varies between bias levels.

In the “IN” configuration we observe the ad’s CTR to be significantly lower compared to that of the first organic result, indicating

Table 4: Learnt feature weights in logistic regression.

| Variable          | Weight Range          | Weight Range          | Weight Range          |
|-------------------|-----------------------|-----------------------|-----------------------|
| \( f_{direct}^i \) | [-0.34, 0.00]         | [0.50, 0.98]          | [-0.80, -0.39]        |
| \( f_{indirect}^i \) | [0.00, 0.02]          | [0.50, 0.89]          | [0.0]                 |
| \( f_{ineffective}^i \) | [-1.77, 0.00]        | [0.21, 0.00]          | [0.62, 1.13]          |
Table 5: t-tests results for the comparisons between the CTR of top ranked entries, \(Ad\) stands for an ad entry, \(o_1\) stands for the \(i\)'th organic entry, \(\checkmark\) for \(p < 0.05\), and \(\times\) for \(p \geq 0.05\).

|  | Y | M | N | IY | IM | IN | DY | DM | DN |
|---|---|---|---|----|----|----|----|----|----|
| \(Ad - o_1\) | \(\times\) | \(\times\) | \(\times\) | \(\checkmark\) | \(\checkmark\) | \(\checkmark\) | \(\checkmark\) |
| \(o_1 - o_2\) | \(\checkmark\) | \(\times\) | \(\checkmark\) | \(\checkmark\) | \(\checkmark\) | \(\times\) | \(\times\) |

that ads are generally ignored by the users. The second organic result’s CTR is also significantly lower than the first organic result’s CTR, similar to its corresponding no-ads configuration ‘N’. For the “IY” and “IM” configurations, however, the CTR of the ads is not significantly different than that of the organic results. Examining Figure 2b and the second row of the table we additionally observe that for the “IY” configuration, the CTR of the first two organic results is also similar.

In these two configurations, we do not see the strong preference towards organic results exhibited in the negative bias configuration and in the direct marketing ads configurations. But we are also not seeing the ranking-biased examination behavior exhibited for the no-ads configurations. Instead, we observe a different pattern in the examination of both sponsored and organic results, with uniform participant attention spread across the top-ranked results, either organic or sponsored. To the best of our knowledge, such a phenomenon was not reported in previous literature.

We believe the reason for these variations stems from the level of contrast between the viewpoints of the organic and sponsored content. When the SERP’s bias level is towards the negative viewpoint, the opinion in the ad is strongly contradicted by the organic results, i.e., the contrast between the ad’s viewpoint and the bias of the organic result is high. The participants can notice this fact, skip the ad, and proceed to click on the organic results. However, when the contrast is not high, participants are more prone to inspect the ad’s content.

Indirect marketing ads are designed for emerging themselves in organic content by mimicking the look and feel of the content surrounding them, with the purpose of reducing users’ negativity towards sponsored content. Our results suggest that this technique is only partly effective and depends on the content of organic results. Previous work showed an interdependence between sponsored and organic results [40, 48]. For indirect marketing ads, we report an interdependence as well.

We shortly revisit direct marketing ads to inspect the second row of Table 5. Here we observe that the CTRs of the first two organic results are not significantly different from each other, in all bias levels. This is opposed to the no-ads configuration in which, for 2 out of 3 configurations, the CTR of the first organic results was significantly higher than that of the second rank. These observations indicate that While participants did not click on the ad, its presence still affected their behavior and caused them to explore top organic results more uniformly. The origin of these behaviors can have multiple explanations related to user cognitive processes during search sessions. We leave these directions as interesting future work.

Our observations lead us to accept hypothesis H1 for direct marketing ads but mostly reject it for indirect marketing ads: Although in one bias configuration (IN), users did skip indirect marketing ads and prefer higher ranked organic results, there is still a non-negligible portion of cases where the users inspected the ads.

### 4.3 Impact on Decision-Making

We now focus on the research question about users’ decision-making under the presence of sponsored content. The main results for this section are summarized in Table 6 according to ad-bias configuration groups, similarly to Table 5.

The posterior bias level is computed using Eq (1) by using the actual CTR observed in our study data: \(P(C|T(r)) = CTR(r)\). As can be seen in Table 1, while the absolute values of the posterior and prior bias levels are slightly different, the order among viewpoint bias remained the same.

The table details the number of responses along with the percentage of responses in parenthesis: “Yes” for a positive response, “Maybe” for an inconclusive response, and “No” for a negative response. Out of 463 responses, only 10 were ‘No Sure’. For clarity of presentation, they were therefore discarded from the analysis.

Our null hypothesis states that sponsored content will not introduce bias to users’ decision making. To test our hypothesis we conducted a set of t-tests in which, for each ad-bias configuration, we compared between all pairs of possible answers to determine the decision bias for that configuration. An ad-biased configuration is considered biased towards one answer class (i.e. “Yes,” “No,” or “Maybe”) if the proportion of responses received for that answer is significantly higher than the alternatives. Bias is considered shared between two answers if their proportions are significantly higher than the third option, but not significantly different from each other. Otherwise, the configuration is considered bias-free.
Table 6: Participants’ response count and percentage value, grouped by ads-bias configurations.

|      | Yes | Maybe | No |
|------|-----|-------|----|
| Y    | 28(0.52) | 14(0.26) | 12(0.22) |
| YI   | 14(0.41) | 15(0.44) | 5(0.15) |
| DY   | 23(0.38) | 28(0.47) | 9(0.15) |
| M    | 12(0.24) | 25(0.50) | 13(0.26) |
| IM   | 25(0.39) | 27(0.42) | 12(0.19) |
| DM   | 11(0.26) | 20(0.47) | 12(0.28) |
| N    | 7(0.14) | 22(0.44) | 21(0.42) |
| IN   | 9(0.17) | 14(0.27) | 29(0.56) |
| DN   | 10(0.22) | 16(0.35) | 20(0.43) |

The values for which the bias tests were significant are highlighted in bold in Table 6. For example, in the ‘Y’ configuration we observe bias towards the “Yes” response, and in the “YI” configuration we observe that the bias is shared between the “Yes” and “Maybe” responses. A post-hoc analysis to determine the power of the t-tests conducted produced power values (1 − β) between 88% to 99% which validates the tests are powered.

The results show that in the no-ads configurations, there is a correlation between the maximal bias level of the SERP and the decision bias exhibited by participants, which aligns with previous work on position bias. This correlation is particularly evident for the “Y” and “M” bias levels, but less noticeable for the “N” bias level. We also observe that for each bias configuration a different user behaviour pattern is exhibited in at least one of the ad configurations.

For indirect marketing ads, we observe ads’ influence in all bias levels, with the direction of the influence depending on the SERP’s bias level. According to our CTR data, in configurations ‘Y’ and “IM” the ads received similar attention to that of the top organic results. This behaviour is reflected in the decisions of users. In both configurations we observe a decrease in the portion of participants that chose the “No” answer, compared to the relating ad free bias level, and as a result a shift in the decision bias from one dominate answer to two: “Yes” and “Maybe”.

In the “IN” configuration, the ad was seldom visited according to our CTR data. Nevertheless, its mere existence seems to have introduced negative bias towards the treatments’ effectiveness. The response bias in this configuration is now towards the “No” response option, as opposed to its corresponding ‘N’ configuration where the bias was shared by the “Maybe” and “No” response options.

Direct marketing ads are also rarely inspected. Nevertheless, their presence can still affect users’ examination of organic results, potentially leading to different responses compared to the corresponding no-ads configurations. Indeed, in the configurations where we observed a different examination behaviour of organic results (“DY” and “DN”) we observe a shift in the dominant bias levels compared to the correlating no-ads configurations. And we no longer observe the correlation between the decision bias and position bias of the SERP, as in the no-ads configurations.

The results provide an affirmative answer to our research question whether sponsored content affects users’ decision making in medical-related search and therefore rejects H2 which hypothesised that ads do not affect users decision making. As our results indicate, sponsored content can introduce either positive or negative bias, depending on the type of the ad as well as on the bias presented by the organic results.

4.4 Impact on Users’ Search Experience

In our study, participants whose SERP contained ads were requested to rate the ads’ effect on their search experience and their decision-making process. The distributions of participants’ responses for both questions are reported in Figure 3. The vast majority of participants reported that the ad did not affect either their search experience or their decision-making. There was no significant difference between the two ads configurations. Eighty participants provided comments about the effect of the ad on their experience. Most of the comments were neutral, noting that they did not mind or notice the ad in the SERPs, and they tend to ignore ads in general. For example: “I ignored the ad that I saw and skipped down to the first non-ad link.”. Some participants expressed dissatisfaction in their comments. For example: “Intrusive and annoying. The world is one big ad.” And some admitted the negative bias that ads invoke in them, for example: “I notice them, but automatically disregard them as unreliable.” To summarize, the participants’ responses and comments indicate that the majority of them perceive themselves as unaffected by ads. However, the results presented in the previous sections suggested that oftentimes this is not the case, as both their examination and decision-making are affected by the sponsored content presented in the SERPs.

5 MITIGATING THE DECISION BIAS

The results presented in the previous section prove that sponsored content can affect users examination behaviour and more importantly their decision making. This phenomenon is most concerning in SERPs where the organic results’ content points towards an inconclusive viewpoint. In this setting, our results show that the existence of indirect marketing ads can persuade users to believe in the effectiveness of a treatment when the organic results do not support that conclusion.
Do Omega Fatty Acids treat ADHD

Polyunsaturated fatty acid supplementation, such as fish oil, is not an FDA-approved treatment for ADHD. The evidence for its effect on ADHD is mixed. "There is little evidence that PUFA supplementation provides any benefit for the symptoms of ADHD in children and adolescents." Obviously further research is needed.

From verified source: https://chadd.org/adhd-weekly/study-shows-omega-3s-benefit-some-children-with-adhd/

Fish Oil Supplements and ADHD - The Evidence is Mixed

Figure 5: SERP with content moderation.

There are ways to counter the bias introduced by sponsored content, by changing the search result page’s presentation. This form of intervention is commonly known as content moderation. Popular content moderation techniques include contextual warning messages that alert users about the reliability of content, interstitial warnings that block certain content and force users to actively select viewing it despite the warning, as well as explicit messages notifying the users that content has been found as inaccurate. Content moderation has been used mainly in social media to control the spread of inaccurate information such as fake news and health related misinformation. Its effectiveness in achieving those goals however has produced mixed results [26, 34, 37, 42].

Content moderation is a complex subject. Applying it wisely while balancing between search engines’ desire to provide reliable information to users and a business model that relies on advertising, is even more complex. In order to properly and thoroughly investigate this issue, a dedicated research study is warranted. Nevertheless, as a proof of concept, we experimented with a few popular content moderation techniques to inspect its possibility to mitigate the bias introduced by sponsored content. We focus our experiments on the "IM" ad-bias configuration, which, as we explained above, is the most concerning setting.

We experimented with two main content moderation approaches: turning negative attention to sponsored content using a warning message and/or pushing it to a lower position in the SERP, and attracting attention to selected organic results using a form of an extended snippet. Figure 5 shows an example of a SERP with two forms of content moderation. The first presented result is an organic result with an extended snippet. The portions of the text most relevant to the query are marked in bold font. In addition we highlight the fact that this result is reliable by adding the title "verified answer" to the snippet and the phrase "From verified source" in red before the URL. The second entry is an indirect marketing ad, with a warning message below it.

We run the additional experiments according to the setting described in Section 3 and analysed the results as in Section 4. The descriptive statistics of the additional experiments are quite similar to those reported in Section 4, we omit the specifics to avoid redundancy.

Figure 4 presents the CTR data in SERPs with content moderation and Table 7 presents participants decisions for the examined content moderation methods. The different content moderation methods are denoted by two letters representing the first two entries in the SERPs in that group: 'M' is an organic result with an inconclusive viewpoint (as in the Section 4), 'X' denotes an extended snippets, 'W' an ad with a warning message, and 'I' an indirect marketing ad without a warning message. For example the SERP in Figure 5 belongs to the "WX" configuration. In all configurations the bias of organic results is towards an inconclusive ("Maybe") viewpoint.

A series of t-tests that were conducted to compare the CTR of the ad to that of the first organic result showed that the CTR of the organic result was significantly higher compared to that of the ads, for all content moderation configurations. Indicating that all content moderation techniques are successful in reducing the CTR of indirect marketing ads.

Inspecting Table 7 however, we observe that not all methods are successful in mitigating the bias introduced by the ads. Despite the warning message’s success in deterring participants from visiting the ads, its presence alone is not sufficient to mitigate the bias towards an "Effective" viewpoint, which was introduced by the title and the snippet of the ad. Encouragingly though, all the methods containing an extended snippet were successful in this task. For these methods we observe that the decision bias correlates with the ranking bias of the SERP as in its corresponding no-ads configuration 

|     | Yes | Maybe | No |
|-----|-----|-------|----|
| MW  | 20(0.39) | 21(0.40) | 11(0.21) |
| XW  | 11(0.17) | 36(0.55) | 18(0.28) |
| IX  | 18(0.3) | 27(0.45) | 15(0.25) |
| XI  | 24(0.31) | 43(0.55) | 11(0.14) |

are not adequate to mitigate the bias introduced by sponsored content.

6 CONCLUSIONS

In this work, we conducted a series of user studies to examine the effect of two types of sponsored content on users’ examination behaviour and decision-making in medical-related web searches.

We found that both types of ads had an effect on users’ examination behaviour as well as on their decisions. The most significant effect was observed for indirect marketing ads, where the contrast between the positive viewpoint presented by the ad and the viewpoint bias presented in the SERP, could introduce either positive bias (for SERPs with low to no contrast) or negative bias (for SERPs with high contrast) towards the ad. In light of these results, we explored the possibility of introducing content moderation techniques to the SERP and showed that such techniques can mitigate the bias introduced by sponsored content.

From our post survey questions we learnt that users dislike ads but perceive themselves as not affected by them. Our results showed that when sponsored content is sophisticated enough it can neutralize this negative bias. However, the organic results’ viewpoint still take precedent over the sponsored content. The accuracy of search results in the health domain is therefore crucial in mitigating bias and misinformation.

Table 7: Participants’ decisions for content moderated SERPs.
REFERENCES

[1] 2012. SEARCH ENGINE USE 2012. https://www pewresearch.org/internet/2012/03/09/main-findings-11/

[2] 2019. Majority of Adults Look Online for Health Information | PEW Research Center. https://www.pewresearch.org/fact-tank/2013/02/01/majority-of-adults-look-online-for-health-information/

[3] Ashish Agarwal, Kartik Hosanagar, and Michael D Smith. 2015. Do organic results help or hurt sponsored search performance? Information Systems Research 26, 4 (2015), 695–713.

[4] Afrah Olayan Alanazi, Mark Sanderson, Zhifeng Bao, and Jaewon Kim. 2020. The impact of ad quality and position on Mobile SERPs. In Proceedings of the 2020 Conf. on Human Information Interaction and Retrieval. 318–322.

[5] Ahmed Allam, Peter Johannes Schulz, and Kent Nakamoto. 2014. The impact of search engine selection and sorting criteria on vaccination beliefs and attitudes: two experiments manipulating Google output. Journal of medical internet research 16, 4 (2014), e100.

[6] Markus Bink, Steven Zimmerman, and David Elsweiler. 2022. Featured Snippets and their Influence on Users’ Credibility Judgements. In ACM SIGIR Conf. on Human Information Interaction and Retrieval. 113–122.

[7] Moira Burke, Anthony Horrobin, Erik Nilsson, and Nicholas Gorman. 2005. High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. ACM Transactions on Computer-Human Interaction (TOCHI) 12, 4 (2005), 423–445.

[8] Georg Buscher, Susan T Dumais, and Edward Cutrell. 2010. The good, the bad, and the random: an eye-tracking study of ad quality in web search. In Proceedings of the 33rd Int’l ACM SIGIR conf. on Research and development in information retrieval. 42–49.

[9] A Cipriani, TA Fukurawa, and C Barbui. 2011. What is a Cochrane review? Epidemiology and psychiatric sciences 20, 3 (2011), 231–233.

[10] Nick Crawwell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An experimental comparison of click position-bias models. In Proceedings of the 2008 Int’l Conf. on web search and data mining. 87–94.

[11] Cristian Danescu-Niculescu-Mizil, Andrei Z Broder, Evgeniy Gabrilovich, Vanja Josifovski, and Bo Pang. 2010. Competing for users’ attention: on the interplay with sponsored results. In Proceedings of the 25th ACM Conf. on Human Information Interaction and Retrieval.

[12] Robert Epstein and Ronald E Robertson. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. Proceedings of the National Academy of Sciences 112, 33 (2015), E4512-E4521.

[13] Morten W Fagerlander and David W Hosmer. 2012. A generalized Hosmer–Lemeshow goodness-of-fit test for multinomial logistic regression models. The Stata Journal 12, 3 (2012), 447–453.

[14] Olivia Fouaidi, Fedz Azzopardi, and Martin Halvey. 2021. Investigating the influence of ads on user search performance, behaviour, and experience during information seeking. In Proceedings of the 2021 Conf. on Human Information Interaction and Retrieval. 107–117.

[15] Claire Gauzente. 2010. The intention to click on sponsored ads—A study of the role of prior knowledge and of consumer profile. Journal of Retailing and Consumer Services 17, 6 (2010), 457–463.

[16] Donna Gillies, John KH Sinn, Sagar S Lad, Matthew J Leach, and Melissa J Ross. 2012. Polyunsaturated fatty acids (PUFA) for attention deficit hyperactivity disorder (ADHD) in children and adolescents. Cochrane Database of Systematic Reviews (2012).

[17] Yessica-Ileana Girardo-Romero, Francisco Muñoz-Leiva, Eulina Higuera-Castillo, Francisco Libanba-Cabaniillas, et al. 2021. Influence of regulatory fit theory on persuasion from google ads: An eye tracking study. Journal of Theoretical and Applied Electronic Commerce Research 16, 5 (2021), 1165–1185.

[18] V Harpin, L Mazzone, JP Raynaud, J Kahle, and P Hodgkins. 2016. Long-term outcomes of ADHD: a systematic review of self-esteem and social function. Journal of attention disorders (2016), 20, 4 (2016), 295–305.

[19] Anit Hashavti, Hongning Wang, Raz Lin, Tamar Stern, and Sarit Kraus. 2021. Understanding and Mitigating Bias in Online Health Search. (2021), 265–274.

[20] Morag Heirs and Mike Emmans Dean. 2007. Homeopathy for attention deficit/hyperactivity disorder or hyperkinetic disorder. Cochrane database of systematic reviews (2007).

[21] Andrew Hershey and Keith J Petrie. 2002. Melatonin for the prevention and treatment of jet lag. Cochrane Database of Systematic Reviews 2 (2002).

[22] Malcolm P Hilton, Eleanor F Zimmermann, and William T Hunt. 2013. Ginkgo biloba for tinnitus. Cochrane Database of Systematic Reviews 3 (2013).

[23] Bernard J Jansen and Marc Reniec. 2005. Examining searcher perceptions of and interactions with sponsored results. In Workshop on Sponsored Search Auctions.

[24] Bernard J Jansen and Marc Reniec. 2006. An examination of searcher’s perceptions of non-sponsored and sponsored links during ecommerce Web search. Journal of the American Society for information science and Technology 57, 14 (2006), 1949–1961.

[25] 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, 9–13.

[26] Ben Kaiser, Jerry Wei, Eli Lucherini, Kevin Lee, J Nathan Matias, and Jonathan Mayer. 2021. Adapting security warnings to counter online disinformation. In 30th USENIX Security Symposium (USENIX Security 21). 1163–1180.

[27] Chuck T Keane, Mave O’Brien, and Barry Smyth. 2008. Are sponsored links effective? Investigating the impact of trust in search engine advertising. ACM Transactions on Management Information Systems (TOMIS) 7, 4 (2017), 1–33.

[28] Anne Oeldorf-Hirsch, Mike Schmierbach, Alyssa Appelman, and Michael P Boyle. 2020. The ineffectiveness of fact-checking labels on news memes and articles. Mass Communication and Society 23, 5 (2020), 682–704.

[29] Mave O’Brien and Mark T Keane. 2006. Modeling result-list searching in the World Wide Web: the role of relevance topologies and trust bias. In Proceedings of the 19th annual conf. of the cognitive science society, Vol. 28. Citeseer, 1881–1886.

[30] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. 2007. In Google we trust: Users’ decisions on rank, position, and relevance. Journal of computer-mediated communication 12, 3 (2007), 801–823.

[31] Omeros Papakriakopoulos and Ellen Goodman. 2022. The Impact of Twitter Labels on Misinformation Spread and User Engagement: Lessons from Trump’s Election Tweets. In Proceedings of the ACM Web Conf. 2022. 2541–2551.

[32] Adrienne Hall Phillips, Ruijiao Yang, and Sousamj Dasmani. 2013. Do ads matter? An exploration of web search behavior, visual hierarchy, and search engine result pages. In 2013 46th Hawaii Int’l Conf. on System Sciences. IEEE, 1563–1568.

[33] Frances A Pogacar, Amira Ghenni, Mark D Smucker, and Charles LA Clarke. 2017. The positive and negative influence of search results on people’s decisions about the efficacy of medical treatments. In Proceedings of the ACM SIGIR Int’l Conf. on Theory of Information Retrieval. 289–296.

[34] Navdeep S Sahni, Harikesh Nair, et al. 2016. Native advertising, sponsorship disclosure and consumer deception: Evidence from mobile search-ad experiments. Journal of Advertising Research 56, 6 (2016), 589–608.

[35] Yan Lu, Michael Chau, and Patrick YK Chau. 2017. Are sponsored links effective? Investigating the impact of trust in search engine advertising. ACM Transactions on Management Information Systems (TOMIS) 7, 4 (2017), 1–33.

[36] Sebastian Schultheiss and Dirk Lewandowski. 2021. How users’ knowledge of search results on people’s decisions about the efficacy of medical treatments. In Proceedings of the ACM SIGIR Int’l Conf. on Theory of Information Retrieval. 289–296.

[37] Ryen White. 2013. Beliefs and biases in web search. In Proceedings of the 36th Int’l ACM SIGIR conf. on Research and development in information retrieval. 3–12.

[38] Ryen White. 2014. Belief dynamics in web search. Journal of the Association for Information Science and Technology 65, 11 (2014), 2165–2178.

[39] Ryen White and Ahmed Hassan. 2014. Content bias in online health search. ACM Transactions on the Web (TWEB) 8, 4 (2014), 1–33.

[40] Shae Sha and Anuanya Ghose. 2010. Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? Marketing science 29, 4 (2010), 602–623.

[41] Shanumi Youn and Seungyoun Kim. 2019. Newsfeed native advertising on Facebook: young millennials’ knowledge, pet peeves, reactance and ad avoidance. Int’l Journal of Advertising 38, 5 (2019), 651–683.

[42] Yinong Yue, Rojan Patel, and Hein Roehrig. 2010. Beyond position bias: Examining result attractiveness as a source of presentation bias in clickthrough data. In Proceedings of the 19th Int’l conf. on World wide web. 1011–1018.

[43] German Zenetti, Tamm H A. Bjornljot, Peter S H. Leeflang, and Daniel Klapper. 2014. Search Engine Advertising Effectiveness in a Multimedia Campaign. Int’l Journal of Electronic Commerce 18, 3 (2014), 7–38.