Impact of Biased Search Results on User Engagement in Web Search

W. R. A. Turk

Delft University of Technology

TU Delft

Impact of Biased Search Results on User Engagement in Web Search

by

W. R. A. Turk

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday July 8, 2021 at 13:00 PM.

Student number:	4599004	
Project duration:	November 27, 2020 – July 8, 2021	
Thesis committee:	Prof. dr. ir. G. J. P. M. Houben,	TU Delft, supervisor
	Dr. L. Chen,	TU Delft
	Dr. U. Gadiraju,	TU Delft, daily supervisor

Abstract

Search engines play an important role in the provision of information. Recently, researchers have raised their concerns about the potential of search engine providers trying to shift the opinion of their users by manipulating search results. But search results might also be manipulated for financial gain. If an increase in User Engagement (UE) can be achieved by manipulating search results, the manipulation of search results can lead to increased advertisement revenue. In this thesis, a $2 \times 3 \times 3$ factorial user study ($n = 376$) investigating the effect of biased Search Engine Results Pages (SERPs) on UE in users with strong opinions on the topic is presented. Different search result ranking biases were used: a search result ranking biased to agree with the topic, one that disagrees with the topic and one that is indifferent towards the topic. One of two different topics is assigned to the user. The bias in the SERP is determined by the condition assigned to the user (i.e., *mitigate*, *control* or *reinforce*) and the viewpoint of the user towards the topic (i.e., *agree* or *disagree*). Furthermore, one out of three queries is presented at the top of the SERP (i.e., a query that *agrees*, *disagrees* or is *impartial* with the topic). It was found that neither mitigating nor reinforcing bias through the SERP affected UE. The exploratory analysis suggests that *confirmation bias* did not guide the results clicked by the users. Users were not refrained from clicking search results that do not align with their viewpoint on the topic. Instead, the exploratory analysis suggests that results are clicked based on their position (*position bias*), regardless of whether bias of the user is *mitigated* or *reinforced*.

Preface

This thesis is the final work I have to complete in order to obtain my Master of Science degree in Computer Science at Delft University of Technology. I have put a lot of work into this thesis and learned a lot while working on it. Completing this thesis would not have been possible without the help of certain people and one organisation. First and foremost, I would like to thank Ujwal Gadiraju and David Maxwell for being my daily supervisors. They helped me a lot in shaping this thesis. Ujwal especially helped me a lot with his insights on the setup of the user studies and the analysis of the results. The logging framework provided by David and his help on using this framework helped me a lot regarding the implementation of the custom search engine. Then, I would like to thank Geert-Jan Houben for being my thesis advisor. Finally, I would like to thank Lydia Chen for being part of the committee for my thesis. I would also like to thank SURF for allowing me to use a virtual machine instance on the national computing facilities. Access to these facilities enabled me to host the custom search engine that was built for this thesis.

W. R. A. Turk
Delft, July 2021

Contents

List of Figures

List of Tables

1

Introduction

In an age where almost all information can be found online, search engines play an important role. Not all information that can be found online can be seen as credible information (e.g., online information in the health domain [73]). Users are uncertain when providing credibility judgements for search results. Even though a lot of non-credible information is present on the internet, most users blindly trust the information that is presented by search engines. For example, an eye tracking study has shown that college students have substantial trust in Google to rank their results optimally w.r.t. the issued query [55]. The trust given to search engines empowers search engine providers to influence society. Outcomes of elections [16] or the users' opinion on the efficacy of medical treatments [56] could be influenced by prioritising selected search results. Allam et al. [1] have shown that search results regarding vaccination can both positively and negatively influence the user. Therefore, in order to avoid unnecessary negative influence, it is important that search results with balanced arguments are shown to the user.

To the users of search engines it is not known whether search engines are willingly manipulating search results. There are many possible explanations as to why search results could be biased. One of these explanations is that search engines would like to shift the opinion of the user on a topic about which the user searches. Concerns have been raised by researchers about search engines trying to shift the opinion of the user [32, 37]. Robertson et al. [60] audited Google Search for a set of political queries. They found significant differences in the composition of politically-related Search Engine Results Pages (SERPs) that were presented to users. However, bias in search results does not have to be intentional. Bias in search results can be caused by *data bias*, *interaction bias* or even *second-order bias*, where implicit biased user feedback is used to optimise web search [6]. Users are also biased themselves, but often do not know that this is the case [59].

Shifting the users' opinion is potentially not the only motivation that exists for manipulating search results. When searching the Web, advertisements are shown in response to the queries issued by users. Because both search results and advertisements look very similar, it has become harder to tell the difference between the two [40]. It has also been shown that users with little knowledge on search advertising are more likely to click on advertisements and less willing to scroll beyond the advertisements to view real search results [62]. Users that are not able to distinguish between search results and advertisements are about twice as likely to click on an advertisement than users that are able to distinguish between search results and advertisements [39]. Furthermore, it has been shown that users behave similarly in advertised search results as in regular search results [4]. So, there is not necessarily a disadvantage for search engine providers to show advertisements to users as if they are normal search results. Letting users interact more with SERPs can thus lead to financial gain.

As mentioned before, there is no evidence that search engine providers do willingly manipulate search results. But, it is important to be aware of any potential dangers. The aim of this research is to investigate whether an incentive for search engine providers exists to manipulate search results. There could be a financial incentive for the search engine provider to increase User Engagement (UE). An increase in UE in both the short and long term could create additional revenue for the search engine provider. If the UE increases, the user for instance could interact with more advertisements, increasing the advertisement revenue. Therefore, if UE can be increased by manipulating search results, this could be a motive for the search engine provider to manipulate search results. Search engine providers could potentially increase UE by exploiting *cognitive dissonance*, where people tend to avoid situations that challenge their beliefs [20]. In web search this is known as a *confirmation bias*, which means that users tend to avoid search results that do not align with their beliefs [37]. Search engine providers could for example show more results to the user of which they know the user agrees with them at the top of the SERP. This could be beneficial for the search engine provider but it could also cause dangers (e.g., for democracy when search is related to politics [9]).

1.1. Research Objectives

The main goal of this thesis is to find out whether an incentive in the form of UE exists for search engine providers to manipulate search results. This is investigated for users with strong opinions, since it is expected that their beliefs will be challenged the most and as a result the largest change in UE will be observed. The overarching research question is:

Overarching Research Question: *How is UE affected in users with strong opinions when SERPs are biased?*

One of the many reasons why it is useful to study whether UE is affected by biased search results is that it is interesting to see whether an incentive exists for search engine providers to reinforce bias. If the search engine provider knows what a user believes and has incentive to reinforce this believe, this could have a negative impact on the user. On the other hand it is interesting to see whether mitigating bias has an impact on the engagement of the user. If mitigating bias has no impact on the engagement of the user, this could be used as an argument to motivate search engine providers to show balanced search results to their users. If it is known that the user believes something that is proven not to be true, more search results that convey the truth can be shown to the user. This all depends on whether UE is affected by showing biased search results. Currently, it is not known how UE is affected when biased search results are shown. Addressing this research question will allow to take action depending on the outcome of this thesis. For example, if it turns out UE is increased when reinforcing bias, this can be used to investigate whether search engines exploit this fact and users can be warned. When it is found that mitigating bias does not affect UE, this can be used as a call for action for search engine providers to show truthful results. It will be investigated whether UE is affected when SERPs are biased by answering the following research questions:

- **RQ1.** What is the impact of *mitigating* bias in users with strong opinions by manipulating the SERP in web search on UE?
- **RQ2.** What is the impact of *reinforcing* bias in users with strong opinions by manipulating the SERP in web search on UE?

To answer these questions, search engines with different biases in the SERP will be presented to users. In the search engine, one of the two selected topics will be shown. The search engines will represent a different bias by changing the composition and ranking of search results. One search result ranking will be biased to agree with the topic, another will be biased to disagree with the topic and the final one will be neutral towards the topic. Search result rankings are biased by changing the positions of search results with particular viewpoints. The search result ranking will be assigned based on the viewpoint of the user and the condition to which the user is assigned. UE will be assessed by recording the behaviour of the participants and recording their responses for a questionnaire on UE items. The data on UE will be analysed to see whether UE is affected by bias in the SERP.

1.2. Contributions & Outline

The following contributions are made through this thesis:

- A $2 \times 3 \times 3$ factorial user study exploring how biased search results affect UE in users with strong opinions is presented.
- All (anonymised) data and code produced and used for this thesis is publicly released.[1] The dataset consists of the search results used in the custom search engine and the corresponding viewpoint annotations. The dataset also consists of the behavioural data recorded and the survey responses from the participants. The code of the custom search engine used in the user study is also published.

We find that UE is not affected by showing biased search results to the user that has a strong opinion on the topic. It is found that the interaction effects between the topic presented to the user and the bias in the search results do not affect UE. This is also the case for the interaction effect between the bias in the search results and the query shown at the top of the SERP, and the interaction effect between the perceived diversity of the search results and the bias in the search results. The exploratory analysis suggests that the lack of *confirmation bias* explains why UE is not affected by biased search results.

This thesis will be outlined as follows. Literature that is relevant for this thesis will be summarised and discussed in Chapter 2. The pilot study, together with the selection of the topics and the creation of the search results will be discussed in Chapter 3. Then, in Chapter 4, the design choices of the custom search engine will be discussed. In Chapter 5, the experimental setup will be discussed. In this chapter, the goal of the final study will be discussed among other topics such as how participants were recruited. Then, in Chapter 6, the results of the final study will be shown. These results will be discussed in Chapter 7. Finally, the conclusions together with recommendations for future work will be given in Chapter 8.

[1] The data can be found at: `https://doi.org/10.4121/14831070`. The code can be found at: `https://doi.org/10.4121/14831079`.

2 Related Work

In this chapter the literature that is relevant for this thesis will be discussed. In Section 2.1, a high level overview of different cognitive biases will be given. Then, in Section 2.2, literature regarding the effect of search results on beliefs and behaviour of the user will be discussed. Thirdly, studies regarding UE in web search will be discussed in Section 2.3. Finally, in Section 2.4, related work regarding critical thinking in web search will be discussed.

2.1. Cognitive Biases in Web Search

Novin and Meyers [50] describe several potential sources of bias related to the way in which information is presented to the searcher. In this study four different kinds of cognitive biases are focused on:

1. *Priming effect* occurs when the choice made by the searcher is influenced by a stimulus [64], a stimulus could be provided by for example the repeated use of a layout.

2. *Anchoring effect* happens when the searcher attaches more value to the first result in the SERP than other results [63].

3. *Framing effect* causes the searcher to make different choices depending on how the information is presented [30].

4. *Availability-heuristic* refers to when a searchers' estimate is influenced by the information that the searcher immediately recalls [63].

The study subjects were asked to rank five different search results. The results of the experiment show that the subjects score Wikipedia articles higher since they believe Wikipedia provides an objective perspective. This bias is also called *domain bias* [28]. Ieong et al. [28] have shown that by just changing the domain of a search result, the preference of the user can be changed 25% of the time. Other bias that exists in web search is *position bias*, which is related to the *anchoring effect*. Keane et al. [33] have shown *position bias* exists by analysing normalized percentages of first clicks in a normal and reversed condition. The highest-ranked items in the normal condition are clicked less often in the reversed condition, while the lowest-ranked items in the normal condition are clicked more often in the reversed condition. The position of a search result might also instigate in the *order effect*, causing users to assign more value to the information that is encountered first [7, 38]. Azzopardi [5] additionally mentions the *exposure effect* and *confirmation bias*. It has been shown that when a user spends more time engaging with a certain viewpoint or stance, the user tends to have a more positive judgement of that viewpoint or stance, this is known as the *exposure effect* [38, 66]. Furthermore, *confirmation bias* causes people to disregard or dismiss information when that information is contradictory to what those people believe [35, 65].

The causes and effects of these cognitive biases have been widely studied. However, it is unknown how these biases affect UE. For example, no studies have looked at the effect of increased exposure of search results with a certain viewpoint in the higher-ranked items on UE.

2.2. Beliefs and Behaviour

Various studies have looked at beliefs and behaviour of users in search engines. Pothirattanachaikul et al. [58] have investigated how collective questions and answers, also known as *People also ask*, impact the users' beliefs and search behaviour. The participants of the study were required to look up information with a customised search engine in order to answer medical yes-no questions. To determine how the users' beliefs changed, it was assessed whether the user changed his/her beliefs while using (or not using) the *People also ask* feature. In order to see how the users' behaviour is impacted, several metrics are recorded:

- **# of Queries:** number of queries issued during the search task.
- **# of Clicks:** number of SERP clicks executed during the search task.
- **# of Evidence:** number of documents selected by the user as supporting their beliefs after the search task.
- **Deepest Document Rank:** rank of the lowest document clicked.
- **Page Dwell Time:** average time spent viewing documents.
- **SERP Dwell Time:** average time spent on the SERP.
- **Task Time Spent:** time spent on the search task.

Other studies looked at how showing biased results impact the users' beliefs and behaviour in the medical domain as well. One study looked at how the opinion and credibility of a document affect the search behaviour and belief dynamics of the user for health-related yes-no questions [57]. The study identified documents that were consistent and inconsistent with the beliefs of the user. For the belief dynamics it was assessed whether the prior beliefs changed from one polar to the opposite polar. For measuring the search behaviour the same metrics as in [58] were used. Additionally, metrics measuring the number of clicks on documents that were consistent/inconsistent with the prior beliefs of the user were used. The results show that users spend more effort when documents that are inconsistent with their prior beliefs are shown. They also tended to change their beliefs when inconsistent results were shown, while they remained their beliefs when documents consistent with their current beliefs were shown. Another work has shown that people favor positive information over negative information when searching for information in the health domain [65]. Allam et al. [1] have studied how the results shown on the SERP affect beliefs regarding vaccination. A custom search engine that manipulated the sites that could be retrieved by the study participants was used. The study participants were told to inform themselves about vaccination. The beliefs of the study participants were recorded by using questionnaires, both before and after the search task. The experiments conducted in the study show that evidence-based medical information can positively impact the beliefs of the user. On the other hand, the results show that documents retrieved by biased search engines can negatively impact the beliefs of the user. It has also been shown that the search results shown to the user can significantly affect their decisions on the efficacy of medical treatments [56]. This confirms that medical-related beliefs are affected by the results that the user gets to see. A follow-up study showed that even when a think-aloud process is used, people are significantly influenced by misinformation [23]. In order to improve search engines for the medical domain, eye patterns can be used with a level of certainty to determine the health literacy of the user [43].

The domain of politics is also a relevant domain for studying the impact of search results on beliefs. An audit of Google Search has shown that significant differences in search results and personalisation of politically-related SERPs exist [60]. Another study has shown that spammers could introduce biased information into the search results of Bing and Yahoo [46]. Epstein and Robertson [16] have studied the impact of search engine manipulation on the outcomes of elections. It was found that it is possible to shift the voting preferences of undecided voters by 20%. Furthermore, it was found that bias in the search ranking can be masked such that the users are not aware of the manipulation. A more recent work aimed to suppress the Search Engine Manipulation Effect (SEME) [17]. The study shows that by introducing bias alerts it is possible to reduce a SEME. Moreover, a SEME can be eliminated entirely by alternating search results with an equal-time rule. Draws et al. [13] have looked at how different search result rankings affect attitude change in web search. Their results indicate that the *order effect* does not cause a SEME. For the *exposure effect*, however, the results indicate that it could contribute to a SEME.

Even though there are many papers that study how biased search results affect beliefs and behaviours of users, no papers were found that look explicitly at how UE is affected. Furthermore, the papers that were found are about biased search results in the medical and political domain. No papers were found that investigated the effect of biased search results for controversial issues outside these domains. For example, in the work by Pothirattanachaikul et al. [57], the opinions might not be very strong since the study requires participants to answer medical yes-no questions. The topic of vaccination in the work by Allam et al. [1] could be considered a controversial topic. But, this work only focuses on how this affects the users' beliefs and not the UE.

2.3. User Engagement

Measuring UE in web search is important in order to specify how the user experiences searching the Web. UE metrics are often used to measure how the user experience changes when introducing a new feature or changing the results in web search. Miroglio et al. [47] have studied how adblocking affects UE within the Web. Firefox browser usage data is used to measure UE in relation to adblocking. Since the study looks at interacting with the Web in general, and not necessarily searching using search engines, not all features used are relevant for measuring UE in web search. The features that are relevant for web search are the length of the session and the total number of searches. Other studies did focus on measuring UE in web search [11, 14, 25, 34, 52]. Some interesting metrics that can be used to measure UE have already been mentioned in the previous subsection. Guo et al. [25] have investigated how user interaction features among other features can be used to predict query performance. The user interaction features quantify how the users interact with the search engine. Some of the features that are included and have not been mentioned before are:

- **SATCount:** number of SERP clicks followed by a dwell time of at least 30 seconds.
- **SATRate:** *SATCount* divided by the number of document clicks.
- **AvgClickPos:** average click position in the SERP.
- **AvgNumClicks:** average number of SERP clicks per query.
- **AbandonmentRate:** fraction of times a query is issued but no result in the SERP is clicked.
- **PaginationRate:** fraction of times the next page with results is requested.

Additional metrics that could be useful are given in [52]:

- **TimeToClick:** average time between issuing the query and the first SERP click.
- **Mouseover:** number of mouseover events.
- **NoMouseover:** number of queries without a mouseover event.
- **Scroll:** number of scroll events.
- **NoScroll:** number of queries without a scroll event.
- **ReformulateQueries:** number of queries that are quickly reformulated (within 30s of issuing the original query).
- **UniqueQueries:** number of queries issued that have not been issued by any other study participant.
- **UniqueQueryTerms:** number of query terms used that have not been issued by any other study participant.
- **UniqueResults:** number of results clicked that have not been clicked by any other study participant.

The aforementioned metrics do not indicate how users actually interact with the SERPs. When measuring UE and attention, eye tracking is a popular method [29, 44, 54, 72]. Eye-tracking equipment is often expensive and therefore it is only possible to do a lab experiment with eye-tracking. However, studies have shown that cursor data can be used as a replacement for eye-tracking data [48, 61]. Guo et al. [26] have introduced the Fine-grained Session Behaviour model that can be used to predict search success. It has been shown that by using a motif-based learning framework for mouse movements, user satisfaction can predicted more accurately [42]. Arapakis and Leiva [2] have used mouse cursor position to predict different levels of user engagement for the knowledge module display in a search engine. There are also multiple studies that have looked at how mouse data can be used to predict engagement with webpages [3, 24].

The studies mentioned previously look at metrics for UE within the session. However, looking at long-term UE is also important. Drutsa et al. [14] have investigated regularity and periodicity of search engine use. In the study it is shown that classes of periodicity exist that can be used as a measure of UE. In another study, UE with websites is measured using the absence time [15]. Due to the nature of this thesis it is not possible to use inter-session metrics to measure UE.

Liu et al. [41] have shown that not all queries in a search session contribute the same towards user satisfaction. The satisfaction of a query is affected by other queries and the user satisfaction of the session correlates the most with the last query (also known as the *recency effect*). This means that just using the metrics previously mentioned might not be enough to determine the user satisfaction of a session. Additionally, the User Engagement Scale (UES) can be used. O'Brien et al. [51] have proposed both a revised long-form and short-form version of the UES. The original UES consists of six factors and 31 items. The original factors are: *focused attention*, *perceived usability*, *aesthethic appeal*, *endurability*, *novelty*, and *felt involvement*. In the revised version, four factors are used. The first three factors are *focused attention*, *perceived usability* and *aesthethic appeal*. The last factor is a new factor called the *reward factor*, which is a combination of two *endurability* items and one *novelty* item. The factors are described as follows:

- *Focused attention*: ”feeling absorbed in the interaction and losing track of time” [51].
- *Perceived usability*: ”negative affect experienced as a result of the interaction and the degree of control and effort expended” [51].
- *Aesthetic appeal*: ”the attractiveness and visual appeal of the interface” [51].
- *Endurability*: ”the overall success of the interaction and users' willingness to recommend an application to others or engage with it in future” [51].
- *Novelty*: ”curiosity and interest in the interactive task” [51].
- *Felt involvement*: ”the sense of being 'drawn in' and having fun” [51].

The UES thus focuses on UE in the short-term (e.g., the *focused attention* subscale) and the long-term (e.g., the *endurability* subscale). In total, the revised long-form and short-form UES consist of 30 and 12 items, respectively.

Xu et al. [68] have investigated the effect of opinions of users on their interactions with search results on the Web. They found that users who strongly agree with a topic issued more clicks and spent significantly more time interacting with the search results. The interaction behaviour in this study was logged using JavaScript and the jQuery library. The other studies mentioned before do not go into great detail on how they implemented logging user interactions needed to measure UE. Recently, a logging framework called LogUI was presented [45]. This logging framework consists of a client side as well as a server side application. Using the framework provides less noisy and incomplete data than other frameworks or libraries. It is for example possible to turn the logging of data off when the user is scrolling through the page.

Many metrics are used to measure UE for web search. Most studies in which UE is measured using online metrics use the same common base set of metrics. Some of the online metrics in this set are *# of Queries*, *# of Clicks* and several dwell time metrics. Apart from this base set of measures, a wide variety of additional measures is used. It is not clear what metrics measure UE well, this also depends on the context of web search in which the metrics are used. Therefore, it is important to include a form of the UES as well. The UES has been proven to reliably measure the subjective UE. The UES short-form makes it easier to use it to measure UE in an experimental setting [51].

2.4. Critical Thinking in Web Search

Users trust in the ability of search engines to optimally rank the search results w.r.t. the issued query [55]. Furthermore, it has been shown that most users are uncertain when assessing the credibility of sources and often diverge from the ground truth [32]. Users' search behaviour and likelihood of verifying information are dependent on their verification attitudes [69]. It has for example been shown that elderly use a set of credibility indicators without knowing what they imply when searching for medical information [72]. A survey has shown that even when students know what sources are credible, they do not always use these sources [8].

Several attempts have been made to improve the critical information seeking process. Yamamoto et al. [71] have introduced a web access literacy scale that can be used to assess the ability of users to retrieve accurate information. This scale can be used to educate people on how to critically search for information. Another study looked at query priming in order to promote critical thinking in web search [70]. Both query auto-completion and query suggestion are used to stimulate critical thinking. The query priming approach did cause the participants to issue more queries and more frequently visit the SERP. But, there is not yet sufficient support to show that query priming leads to sustained critical thinking in web search.

The fact users trust in the ability of search engines to optimally rank search results w.r.t. the issued query shows that it is important to investigate whether incentive exists for search engine providers to do otherwise. If this is the case, users can be made aware to stimulate critical thinking. Tools could then be developed to aid users in this process by stimulating critically assessing search results.

3 Reinforcing & Mitigating Users Viewpoints

In this chapter the procedure of the pilot study will be discussed. In the pilot study it will be evaluated how UE changes over different search result rankings that *reinforce* or *mitigate* users' viewpoints. **The purpose of the pilot study is to find the search result ranking that will be used in the search engine for the final study**. That is, to find the search result ranking for which UE is higher when reinforcing bias and lower when mitigating bias. This is because it is expected that reinforcing bias will lead to an increase in UE, while mitigating bias will lead to a decrease in UE. All data related to the pilot study is openly available online.[1] First, the topics for which search results will be shown are selected in Section 3.1. Then, search results are retrieved and search results are created to be used in the search engine for the pilot and final study. This is explained in Section 3.2. Finally, the setup of the pilot study and the results will be discussed in Section 3.3.

3.1. Topic Selection

Before search result rankings can be evaluated, the debated topics for which SERPs will be shown need to be selected. Since the effect of biased SERPs on UE in users with strong opinions is analysed, topics for which most users have strong opinions will be selected. In order to select topics, the dataset of the preliminary study by Draws et al. [13] was used. This dataset consists of opinions from 100 people on 18 topics. The topics were retrieved from *ProCon*[2], a website that contains information on controversial issues. Responses were recorded on a 7-point Likert scale. In total, seven people failed at least one of the two attention checks and they were also excluded from this analysis.

Topics were selected based on two criteria. The first criterion is that the opinions about the topics are not skewed towards the positive or negative side. In order to achieve this, a one-sample Wilcoxon test was conducted. The responses are first transformed to numerical responses (i.e, between -3 and 3, corresponding to "Strongly disagree" and "Strongly agree", respectively). The Bonferroni correction was applied to correct for multiple testing. The new p-value that was used to test for significance is equal to $\frac{0.05}{18} \approx 0.0027$. Significant results suggest that the mean attitude is not equal to 0. Therefore, the goal is to find topics for which the results are non-significant. Secondly, the goal was to find topics for which people have strong opinions. Strong opinions include the following responses: "Strongly disagree", "Disagree", "Agree" and "Strongly Agree". The topics were then sorted based on the number of strong opinions, and the topics with the most strong opinions were selected. It was chosen to select the first two topics, since for those topics the fraction of strong opinions is slightly larger than 0.5. For the remaining topics the fraction of strong opinions is much less than 0.5. The topics that were selected are:

1. *Should more people become vegetarian?*

[1] The data can be found at `https://doi.org/10.4121/14831070`.

[2] `https://procon.org`

2. *Should students have to wear school uniforms?*

From now on the first topic will be referenced as the topic about *school uniforms* and the second topic as the topic about *vegetarianism*.

3.2. Search Results

Search results for the search engine are retrieved using the *Bing Custom Search API*.[3] For both topics 14 different queries are issued. The queries are deduced from the queries used by Draws et al. [13]. The queries can be found in Table 3.1. For each query the top 50 search results are collected. The top 50 search results for each query are combined into one large pool. The pools are processed as follows:

1. Each search result in the pool must be unique. This is ensured by using the URL of the search result. The *school uniform* pool consists of 275 unique search results, while the *vegetarianism* pool consists of 333 unique search results.

2. When no duplicate search results are present within the search result pool, non-relevant search results are removed. This is done by manually visiting the URL and analysing whether the search result is relevant. Search results are relevant when they are related to the topic and provide information that can be used to learn more about the topic (e.g., the search result is not an advertisement). After filtering out non-relevant search results, 87 search results were left in the *school uniform* results pool, and 106 search results were left in the *vegetarianism* results pool.

3. Based on the unique relevant search results, the documents used within the search engine will be created. Each web page can be split into one or multiple documents that each has a single viewpoint. For example, a web page with both pros and cons of becoming vegetarian is split into two documents, one containing the pros and one containing the cons of becoming vegetarian. Sometimes it is too complicated to create one or multiple documents that do not contain multiple viewpoints. When this is the case the search result is discarded and no documents are created for the custom search engine. This step results in a total of 104 documents for the *school uniform* topic (52 pro and 52 con) and 114 documents for the *vegetarianism* topic (64 pro and 48 con).

4. The documents are annotated on their viewpoint. Because of the procedure with which the documents are created, it is clear which viewpoint exists in which document. The viewpoint annotation is given using a 5-point Likert scale. Positive annotations are given to documents that agree with the topic, while negative annotations are given to documents that disagree with the topic. The extreme annotations (-2 and 2) are given to documents of better quality. Documents of better quality are documents that do not just list pros or cons but instead look more like a real search result (e.g., an essay).

After processing the search results and creating the documents they are grouped based on predefined queries. The predefined queries are shown in Table 3.2. Results are ranked using cosine similarity between the query and the documents. The search result rankings differ in bias and the ranking algorithm used.

3.3. Pilot Study

3.3.1. Setup

The pilot study had been approved by the Human Research Committee of TU Delft before running the pilot study. Before taking part in the pilot study the participants are asked for their consent. Then, the participant will be asked to indicate his/her viewpoint towards two statements. The statements used are: *"Students should have to wear schooluniforms."* and *"More people should have to become vegetarian."*. Participants that do not have a strong viewpoint towards any of the statements will be removed from the study. Participants that have a strong viewpoint towards at least one of the two topics will get to see the search engines. The search engines will consist of queries related to a topic for which the participant has a strong viewpoint. If the participant has a strong viewpoint towards both statements,

[3] https://www.microsoft.com/en-us/bing/apis/bing-custom-search-api

Topic	
Vegetarianism	*School Uniforms*
should more people become vergetarian?	should students have to wear school uniforms?
should more people become vergetarian? pros	should students have to wear school uniforms? pros
should more people become vergetarian? cons	should students have to wear school uniforms? cons
arguments opposing becoming vegetarian	arguments opposing wearing school uniforms
arguments supporting becoming vegetarian	arguments supporting wearing school uniforms
opinions opposing becoming vegetarian	opinions opposing wearing school uniforms
opinions supporting becoming vegetarian	opinions supporting wearing school uniforms
becoming vegetarian arguments	wearing school uniforms arguments
becoming vegetarian cons	wearing school uniforms cons
becoming vegetarian opinions	wearing school uniforms opinions
becoming vegetarian pros	wearing school uniforms pros
becoming vegetarian pros and cons	wearing school uniforms pros and cons
why is becoming vegetarian bad?	why are school uniforms bad?
why is becoming vegetarian good?	why are school uniforms good?

Table 3.1: Queries issued to the *Bing Custom Search API*. Queries are deduced from the queries used by Draws et al. [13].

the topic will be randomly chosen. It is also randomly chosen whether the viewpoint is *mitigated* or *reinforced*. For example, if the participant strongly disagrees that students should have to wear school uniforms and it is chosen to mitigate the bias, then the results will be biased in favour of wearing school uniforms. Participants are incentivised by giving them a search task, which is shown below:

”Imagine the following scenario. Your government needs to decide on future policies for several debated topics. Since these topics are highly debated, they ask the people about their informed opinion. You have been selected to give your informed opinion regarding <topic>. In order to give your informed opinion you have decided to use the search engines that you will be provided with in the following questions.”

After deciding what type of search results will be shown to the participant, two different search engines will be shown in a randomised order. The cosine similarities between the documents and the queries are calculated. The queries used are listed in Table 3.2. The queries are shown in a randomised order. This is done to avoid any biases introduced by the order in which the queries are shown.

The difference between the first and the second search engine is the ranking of the search results. The search results will be ranked based on the viewpoints of the documents and the cosine similarity with the query. The search results are first ordered on cosine similarity in descending order. The different search result rankings are shown in Table 3.3. The viewpoints shown are for the example mentioned above where the participant strongly disagrees that students should have to wear school uniforms and the bias is mitigated. The search results indicated with an ‘s’ viewpoint are populated after the search results with a numerical viewpoint. These positions will be filled with search results that have the highest cosine similarity with the query and have not yet been included in the SERP. As a result, positions 6-10 and 15-20 in the second search engine can have any viewpoint. For the first search engine all viewpoints are predetermined. For example, the search result at rank one is the document with viewpoint ’2’ that has the highest cosine similarity with the query. When the SERPs need to be biased to disagree with the topic instead, the sign of the numerical viewpoints will be flipped.

In the survey, together with the link to the search engines, questions are presented to analyse how the participants experience the search process and perceive the diversity of the results. The questions asked can be found in sections 5.2.1 and 5.2.3. While the participant is searching, the UE metrics mentioned in Section 5.2.2 will also be recorded. Recording these metrics will function mostly as a test for the final study and will not be used to analyse the difference in UE between the different search result rankings in the pilot study.

Particpants were recruited via *Prolific*[4]. Prolific is a crowdsourcing platform targeted at scientific research. It is common practice to recruit participants from crowdsourcing platforms for interactive information retrieval studies [22, 31, 67]. Another crowdsourcing platform that is often used for scientific

[4] https://www.prolific.co/

Topic	
Vegetarianism	*School Uniforms*
Pros and cons of becoming vegetarian	Pros and cons of wearing school uniforms
The advantages of becoming vegetarian	The advantages of wearing school uniforms
The disadvantages of becoming vegetarian	The disadvantages of wearing school uniforms

Table 3.2: Predefined queries for each topic in the pilot study.

Search result position	*Viewpoint Search Engine 1*	*Viewpoint Search Engine 2*
1	2	2
2	2	2
3	1	1
4	1	1
5	1	1
6	-1	s
7	-1	s
8	-1	s
9	-2	s
10	-2	s
11	2	2
12	2	2
13	1	1
14	1	1
15	1	1
16	-1	s
17	-1	s
18	-1	s
19	-2	s
20	-2	s

Table 3.3: Document viewpoint rankings in the pilot study. The rankings are based on a participant that strongly disagrees with the topic for which the bias will be mitigated. Positive viewpoint annotations are given to documents that agree with the topic, while negative viewpoint annotations are used for documents that disagree with the topic. Search result positions with an ‘s’ are filled solely based on cosine similarity and the documents in these positions can therefore have any viewpoint annotation.

	Condition	
Topic	*Reinforce*	*Mitigate*
Vegetarianism	11	14
School Uniforms	16	16
Total	27	30

Table 3.4: Distribution of the participants in the pilot study over the topics and conditions, where possible the group sizes were balanced.

	p	
Condition	*UES*	*Perceived Result Diversity Score*
Mitigate	0.069	0.503
Reinforce	0.819	0.040

Table 3.5: p-values for the Wilcoxon signed-rank test across different scales and conditions. Only the difference in the *perceived result diversity score* in the *reinforce* condition is statistically significant.

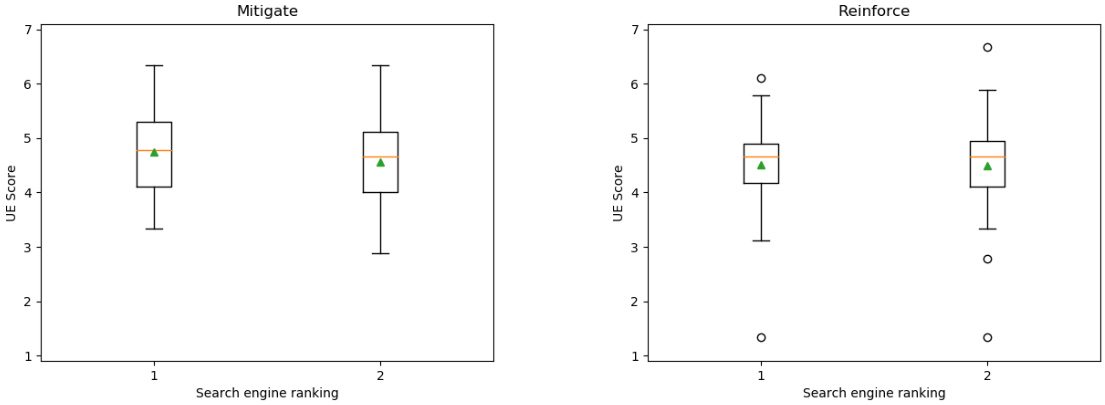
research is Amazon Mechanical Turk (*MTurk*) [10]. Prolific is said to be superior to *MTurk* for sound scientific research in terms of usability [53].

3.3.2. Results

In total 80 participants were recruited. There were 23 participants that did not have a strong opinion on any of the topics. As a result, 57 participants interacted with the search engines. The distribution of the participants over the topics and conditions is shown in Table 3.4. Where possible the participants were distributed equally over the topics and conditions. In order to compare the different SERPs, the Wilcoxon signed-rank test is used. The Wilcoxon signed-rank test is used for both the UES short form and the perceived diversity scale. The ”Aesthetic Appeal” subscale (items 6-9) is (are) removed from the UES short form since this subscale is not related to the search results shown. The null hypothesis of the test is that the median of the differences is zero. If for any of the differences between the SERPs the null hypothesis could be rejected, then this shows that the SERPs are significantly different.

The results for the Wilcoxon signed-rank tests are shown in Table 3.5. At a level of $\alpha = 0.05$, only the difference in the perceived result diversity score for the *reinforce* condition is significant. The boxplots for the UES and the perceived result diversity scores are shown in figures 3.1 and 3.2, respectively. As can be seen in Figure 3.2b, the perceived result diversity score for the second SERP in the reinforce condition is larger than the perceived diversity score for the first SERP in the reinforce condition. Combining this with the fact that the differences are significant, indicates that the second SERP shows more realistic search results. The perceived diversity scale scores in the *mitigate* condition are non-significant ($p = 0.503$). The UES scores in both conditions are similar for the different SERPs (Figure 3.1). However, the scores for the second SERP in the mitigate condition do seem to turn out a little bit lower. But, since the differences are non-significant no conclusions can be drawn from this.

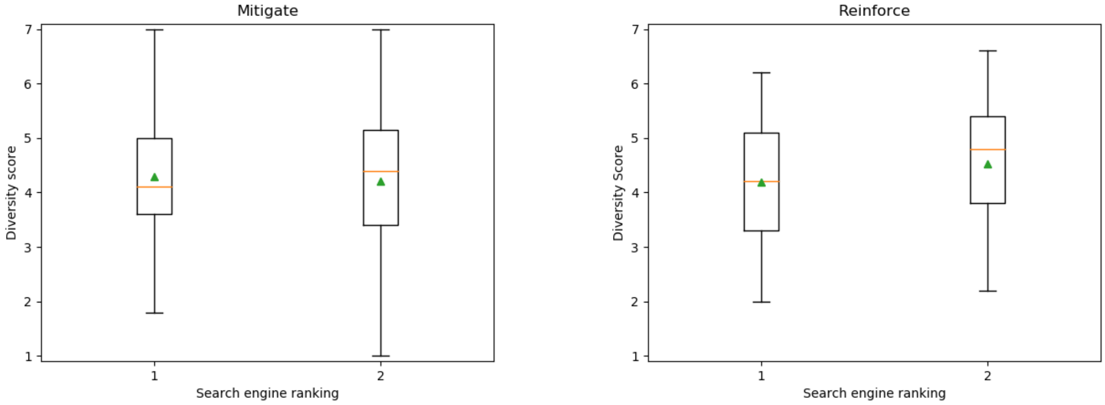
The goal is to have a SERP that is as realistic as possible and a SERP that lowers the UE in the *mitigate* condition, while increasing UE in the *reinforce* condition. Considering the results found, the second search result ranking most suited. This means the search result ranking in which not all viewpoints of the search results are fixed will be used in the final study.



(a) The *mitigate* condition.

(b) The *reinforce* condition.

Figure 3.1: UES scores over different search result rankings for the participants of the pilot study for the *mitigate* condition (3.1a) and the *reinforce* condition (3.1b). There are no statistically significant differences between the first and second search result rankings ($p = 0.069$ and $p = 0.819$ for the *mitigate* and *reinforce* conditions, respectively).



(a) The *mitigate* condition.

(b) The *reinforce* condition.

Figure 3.2: Perceived result diversity scores over different search result rankings for the participants of the pilot study for the *mitigate* condition (3.2a) and the *reinforce* condition (3.2b). In the *mitigate* condition there was no statistically significant difference in perceived diversity score. In the *reinforce* condition the perceived diversity scores for the first and second search result ranking are statistically significantly different ($p = 0.04$). The second search result ranking is thus perceived as more realistic than the first search result ranking by the participants of the pilot study for who the bias is reinforced.

4

Custom Search Engine Design

In this chapter the design and implementation of the search engine used for the pilot and final study will be discussed. First, the URL parameters that are used in the search engine will be discussed in Section 4.1. Then, in Section 4.2, it will be explained how the search results are stored, ranked and loaded. In Section 4.3, the logging of user interactions will be discussed. Finally, in Section 4.4, the user interface of the search engine will be shown and discussed. The code of the custom search engine is made publicly available.[1]

4.1. URL Parameters

In the implementation, URL parameters are used to configure the search engine. The URL parameters consist of information that is related to the setup of the study. Since it is critical that users do not know these parameters, all URL parameters, except the Prolific PID, are encoded using Base64 encoding. The Base64 encoding is non-human-readable format and therefore this will avoid any bias that could be caused by the user knowing the URL parameters. The following URL parameters are used in the implementation of the search engine:

- **Prolific PID.** The Prolific PID is provided such that it is possible to link the data from the survey with the interaction logs that are logged with the interaction logging framework.

- **Topic.** The topic is used to identify which query, or in the case of the pilot study, which queries, need to be shown to the participant.

- **Ranking.** The ranking specifies how the search results should be ranked.

- **Search engine variation (pilot study only).** This URL parameter is used to distinguish between the two different search engines that could be shown to the participant in the pilot study.

- **Query permutation (pilot study only).** The query permutation parameter specifies in what order the queries should be shown to the user in the pilot study.

- **Query (final study only).** The query specifies which query will be shown to the participant in the final study.

4.2. Search Results

The search results play an important role in the studies. In this section it will be described how the search results are stored (Section 4.2.1), how they are ranked in the SERP (Section 4.2.2) and loaded into the search interface (Section 4.2.3).

[1] Code can be found at: `https://doi.org/10.4121/14831079`.

4.2.1. Storage

Search results are stored in Markdown format on the server side. The Markdown format is used since it allows to show search results that look like realistic web pages (i.e., more realistic than plain text). The search results are stored based on the internal id of the webpage from which the search result has been extracted and a number that represents the document within that webpage. For example the second search result that has been extracted from the webpage with id 37 is stored as ”37-2.md”. The first line of a search result file is always the title of the webpage by design. The lines below the first line are the content of the webpage.

4.2.2. Ranking

Multiple search result rankings exist for each query. The search result ranking that is shown to the user is determined by the URL parameter. The different search result rankings are stored in JSON format. For each topic a separate JSON file exists. The JSON file consists of an array of dictionary object. Each dictionary object contains the query itself and three search result rankings. One search result ranking that agrees, one that disagrees and another one that is neutral with the topic. A search result ranking is represented by an array of filenames, which can be used to read the contents of a file.

4.2.3. Loading

Search results are loaded into the interface using JavaScript. When the webpage is loaded, the required search results are read. It is determined what search results are required by using the previously mentioned JSON files that are identified using the URL parameters. The plain text versions of these files are stored and the files are processed using markedjs.[2] The Markdown files are converted to HTML using markedjs such that they can be displayed as a more realistic webpage. The plain text is stored in order to be able to show the title and a small preview of the search result on the SERP.

4.3. Logging User Interactions

The user search interactions are logged using a client- and server-side implementation of the LogUI framework [45].

4.3.1. Client

On the client side the configuration object is specified. In the configuration object parameters like the endpoint, tracking configuration and browser events that need to be tracked are specified. For detailed information on the configuration object one can take a look at the Configuration Object page in the Quick Start Guide of the LogUI Client[3]. In total, nine different search interactions are being tracked. The search interactions can be split up into two different types of events; browser events and within-page events.

There are two browser events that are being tracked. The first browser event that is tracked is the *Page Focus event*. The *Page Focus event* is an event that can be used to determine whether the user is interacting with the webpage or not. For example, if the user minimises the browser or switches to another tab. When this event is fired, a boolean is used to indicate whether the browser of the user is focused on the search engine. Furthermore, the time of each event is indicated and this can be used to calculate the total active time in the search engine. Another browser event that is logged is the cursor movements. The cursor movements are tracked at an interval of 100ms and x and y coordinates in the browser's viewport, the user's monitor and in relation to the entire page are supplied. Furthermore, the option to block browser events while scrolling is set to true. Not setting this feature to true can lead to erroneous search interaction logs.

A total of 7 different within-page event types are logged. Five of these events are click-based, while the other two are hover-based. The following events are logged:

- **A click on the close button of a search result.** This event is fired when the ”click” event is fired on the close button at the upper right corner of a search result. This event is logged using the name *CLICK_CLOSE*.

[2] https://marked.js.org/

[3] https://github.com/logui-framework/client/wiki/Configuration-Object

- **A click anywhere on the screen.** A click anywhere on the screen is recorded using the name *CLICK_ANY*. If another element is clicked for which clicks are already recorded, the *CLICK_ANY* event is not fired. For example, if the close button at the upper right corner of a search result is clicked, only the *CLICK_CLOSE* event will be recorded.
- **A click on a search result.** Clicks on search results to open the corresponding webpage are labelled as *CLICK_RESULT*.
- **A click on the pagination element.** Clicks to go to the first or second page are recorded under the name of *CLICK-PAGE-1* and *CLICK-PAGE-2*, respectively.
- **Hovering over a search result.** For each search result it is logged when the user hovers over a search result. Two separate events are logged for this, one when the user starts hovering over a search result (*RESULT_HOVER_IN*) and one when the user stops hovering over a search result (*RESULT_HOVER_OUT*).

For the *CLICK_RESULT*, *RESULT_HOVER_IN* and *RESULT_HOVER_OUT* events a metadata sourcer is used. Metadata sourcers can be used to extract additional information from the element and this information is included in the event that is logged. For all three events the additional information that is logged is the rank of the result to which the event corresponds. The rank of a search result is an element attribute and is included in the log as *resultRank*.

The application specific data is also declared in the configuration object. Application specific data can be used to link the data of the questionnaire to the interaction logs. The *Prolific PID* is specified in the application specific data to achieve this goal. The other URL parameters are also specified in the application specific data for ease of use. This means that in the pilot study among other URL parameters the *search engine variation* and *query permutation* parameters are specified, while in the final study the *query* URL parameter is specified in the application specific data.

4.3.2. Server

The server-side LogUI code is run on the same virtual machine as the search engine is hosted on. The server-side is set up using the First Run Guide.[4] Separate flights are created for the pilot study and final study in order to be able to separate the interactions.

4.4. User Interface

4.4.1. Implementation

PHP is used to implement the user interface of the search engine. PHP is a general-purpose scripting language. This language enables to hide logic from users. Furthermore, it allows to dynamically generate HTML elements based on the data that must be shown in the search interface. This is for example used to load the queries into the drop-down list presented at the top of the page of the pilot study. To do this, first, URL parameters are used to load the corresponding file. Then, the contents of the file are read and HTML elements are created for each query. This functionality is also used to create the correct number of HTML elements to show the search results on the SERP.

4.4.2. Search Functionality

The search functionality differs for the pilot and final study. The search functionality for the pilot study is shown in Figure 4.1. For the final study the search functionality is shown in Figure 4.2. In the pilot study the participants are allowed to issue three different queries. The queries in the pilot study are presented as a drop-down list. The user can select the query that he/she would like to issue and press search afterwards. When the search button is clicked, the results corresponding to the currently selected query are shown. Before the search results are shown, a spinner will be presented to the user. The spinner is shown for a random amount of time between 500 and 1000 milliseconds. In theory the search results could be changed instantly. The spinner is presented to the user to create a realistic search experience. By presenting the spinner to the user it becomes apparent that the results shown on the SERP have changed. The spinner is shown in Figure 4.3. In the final study the participants are assigned one query that cannot be changed. The search functionality is therefore not used in the final

[4] `https://github.com/logui-framework/server/wiki/First-Run-Guide`

study. However, when the participant first enters the webpage, the spinner will be presented before the search results are shown. As can be seen in Figure 4.2, the text input and the search button have both been disabled for the final study.

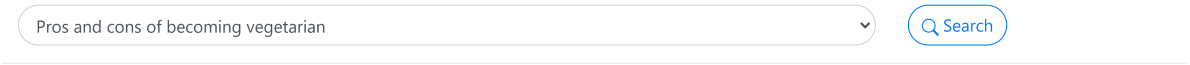


Figure 4.1: Search functionality in the pilot study.

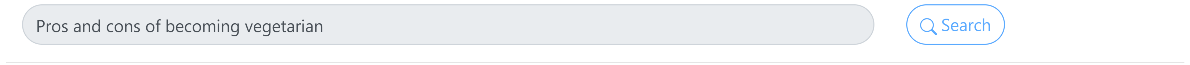


Figure 4.2: Search functionality in the final study.

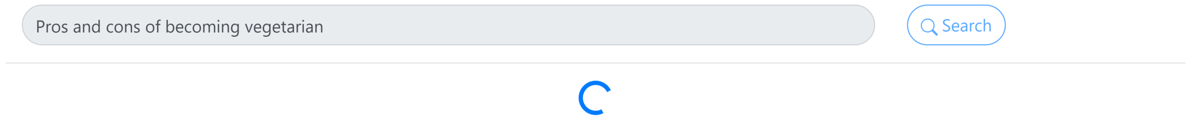


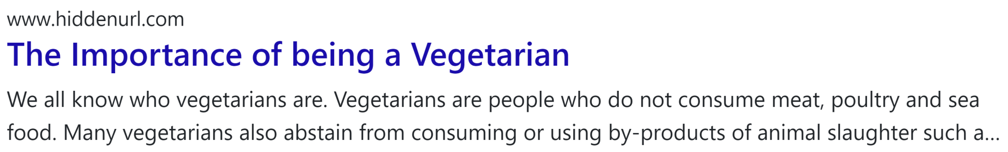
Figure 4.3: The spinner that is shown when search results are being loaded.

4.4.3. Search Results

Search results are shown on the SERP. For each search result that is displayed, the title is shown and below the title a small preview of the search result is shown. Each search result is preceded by the URL *”www.hiddenurl.com”*. This URL is used to avoid any domain bias [28], but still makes the SERP look more realistic. The preview of the search results is the first part of the search result (excluding the title). An example of a search result on the SERP is shown in Figure 4.4. The HTML elements are created using PHP and the content of the HTML elements is set using JavaScript. When a search result is clicked, the search result is opened in an overlay, as shown in Figure 4.5. JavaScript is used to make the overlay visible. The overlay is used for different search results and therefore JavaScript is used to change the content that is displayed in the overlay. In order to process the search result file to HTML the markedjs library is used.

4.4.4. Pagination

For each query a total of 20 search results are shown in the SERP. The search results are not all shown on the same page. Each page of search results contains ten search results. Ten is the most conventional number of search results per page used by most search engines. In order to facilitate the change of search result pages, a pagination item is shown. This pagination item can be found at the bottom of the page. Figure 4.6 shows the pagination item when the first page of results is shown to the user. The currently selected page is highlighted in blue.

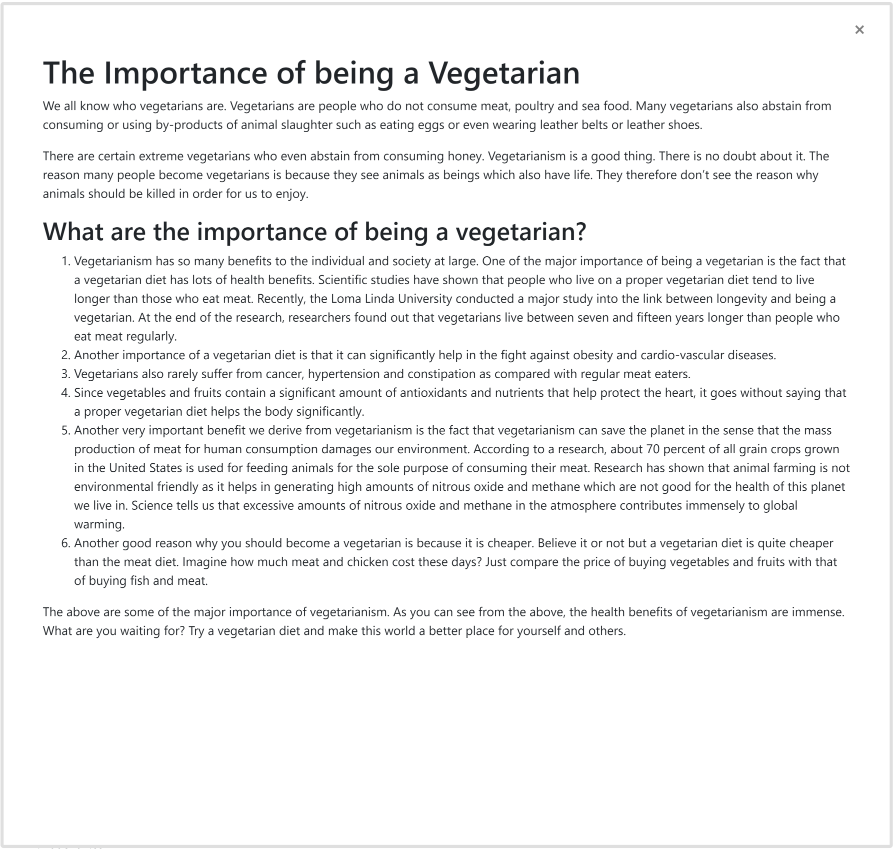


www.hiddenurl.com

The Importance of being a Vegetarian

We all know who vegetarians are. Vegetarians are people who do not consume meat, poultry and sea food. Many vegetarians also abstain from consuming or using by-products of animal slaughter such a...

Figure 4.4: Example of how a search result is shown in the SERP.



The Importance of being a Vegetarian

We all know who vegetarians are. Vegetarians are people who do not consume meat, poultry and sea food. Many vegetarians also abstain from consuming or using by-products of animal slaughter such as eating eggs or even wearing leather belts or leather shoes.

There are certain extreme vegetarians who even abstain from consuming honey. Vegetarianism is a good thing. There is no doubt about it. The reason many people become vegetarians is because they see animals as beings which also have life. They therefore don't see the reason why animals should be killed in order for us to enjoy.

What are the importance of being a vegetarian?

1. Vegetarianism has so many benefits to the individual and society at large. One of the major importance of being a vegetarian is the fact that a vegetarian diet has lots of health benefits. Scientific studies have shown that people who live on a proper vegetarian diet tend to live longer than those who eat meat. Recently, the Loma Linda University conducted a major study into the link between longevity and being a vegetarian. At the end of the research, researchers found out that vegetarians live between seven and fifteen years longer than people who eat meat regularly.
2. Another importance of a vegetarian diet is that it can significantly help in the fight against obesity and cardio-vascular diseases.
3. Vegetarians also rarely suffer from cancer, hypertension and constipation as compared with regular meat eaters.
4. Since vegetables and fruits contain a significant amount of antioxidants and nutrients that help protect the heart, it goes without saying that a proper vegetarian diet helps the body significantly.
5. Another very important benefit we derive from vegetarianism is the fact that vegetarianism can save the planet in the sense that the mass production of meat for human consumption damages our environment. According to a research, about 70 percent of all grain crops grown in the United States is used for feeding animals for the sole purpose of consuming their meat. Research has shown that animal farming is not environmental friendly as it helps in generating high amounts of nitrous oxide and methane which are not good for the health of this planet we live in. Science tells us that excessive amounts of nitrous oxide and methane in the atmosphere contributes immensely to global warming.
6. Another good reason why you should become a vegetarian is because it is cheaper. Believe it or not but a vegetarian diet is quite cheaper than the meat diet. Imagine how much meat and chicken cost these days? Just compare the price of buying vegetables and fruits with that of buying fish and meat.

The above are some of the major importance of vegetarianism. As you can see from the above, the health benefits of vegetarianism are immense. What are you waiting for? Try a vegetarian diet and make this world a better place for yourself and others.

Figure 4.5: Example of how a search result is shown in the overlay.

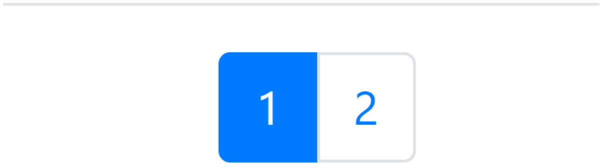


Figure 4.6: Pagination item when the first page of search results is shown.

5
Experimental Setup

In this chapter, the materials, procedure, and participants of the user study will be described. The goal of the final study is described in Section 5.1. Secondly, in Section 5.2, the measurements that are used during the user study are described. In Section 5.3, the variables of the $2 \times 3 \times 3$ factorial user study will be discussed. Then, the steps that the participants will undergo are explained in detail in Section 5.4. In Section 5.5 the statistical analyses that will be used are discussed. Finally, in Section 5.6, it will be described how participants were recruited.

5.1. Goal

The goal of the user study is to answer the research questions. The research questions will be answered by testing several hypotheses. Figure 5.1 shows the conceptual model that represents the hypotheses. The hypotheses that are represented using the conceptual model are:

H1. The composition and ranking of SERPs affects UE in users with strong attitudes towards the topic.

H2. The topic moderates the effect of the composition of the SERP on UE.

H3. The alignment of the query stance with the viewpoint of the participant moderates the effect of the composition of the SERP on UE.

H4. Perceived result diversity moderates the effect of search result composition on UE.

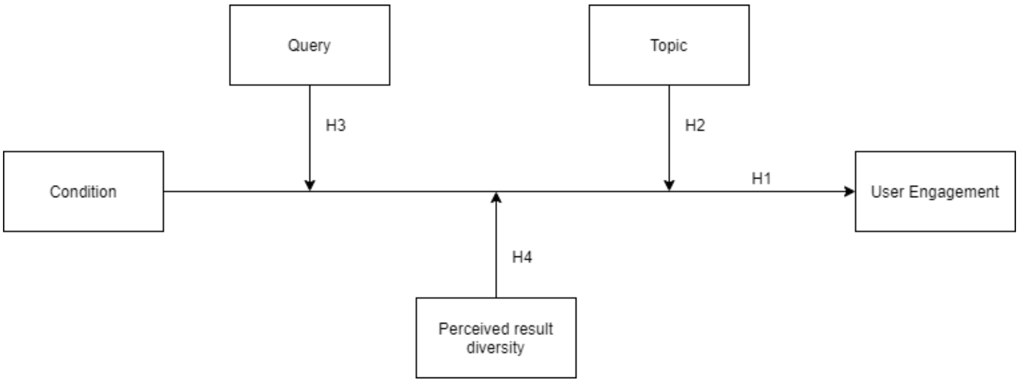


Figure 5.1: Graphical representation of the hypotheses that are tested.

It is expected that the composition and ranking of a SERP affects UE (**H1**) due to *cognitive dissonance*. *Cognitive dissonance* is mental discomfort that is experienced when beliefs are challenged, and people will try to avoid these situations [12, 21, 27]. As a result of mental discomfort it is expected that users will be less engaged. It is expected that not only the search results contribute towards *cognitive dissonance*, but also the query. The expectation is that users will feel like their beliefs are challenged

Topic	
Should more people become vegetarian?	*Should students have to wear school uniforms?*
Pros and cons of becoming vegetarian	Pros and cons of wearing school uniforms
The advantages of becoming vegetarian	The advantages of wearing school uniforms
The disadvantages of becoming vegetarian	The disadvantages of wearing school uniforms

Table 5.1: Queries that can be assigned to the participants for each topic.

less when a query that agrees with their viewpoint is shown and more when a query is shown that disagrees with their viewpoint. This is why **H3** is included. For some topics people have stronger opinions than for others. Therefore, it is expected that the topic will moderate the effect of the composition and ranking of the SERP on UE (**H2**). Finally, it is predicted that the effect of condition on UE will be moderated by the perceived result diversity (**H4**). The rationale behind this hypothesis is that people might feel like their beliefs are challenged less or more depending on how diverse they perceive the results to be.

In order to test the different hypotheses, a participant is assigned to one of the 18 different groups. The groups differ in the bias in the SERP, the query for which results are displayed and the topic for which the query and results are shown. The selection of the topics is discussed in detail in Section 3.1. The following topics are selected:

1. *Should more people become vegetarian?*
2. *Should students have to wear school uniforms?*

One out of three different queries is selected. An overview of the available queries for each topic is shown in Figure 5.1. Three different bias conditions exist: *mitigate*, *reinforce* and *neutral*. Together with the opinion of the participant towards the topic that is assigned, the condition will determine which viewpoint towards the topic the document in each search result position must have. The different search result rankings that exist are shown in Table 5.2. In Table 5.3 it is shown how the different search result rankings are assigned to the participant based on the opinion of the participant towards the topic. Examples of SERPs with different search result rankings are shown in Appendix A. The SERPs that are biased to agree, disagree and not biased are shown in figures A.1, A.2 and A.3, respectively.

5.2. Measures

In this section the measures used for the user study are described. Most have been mentioned in the Related Works (Chapter 2). In this section it will be described in more detail what will be used.

5.2.1. User Engagement Scale

In order to subjectively measure the UE after the experiment, the UES short-form will be used [51]. The UES short-form consists of 12 items that are divided over four subscales: *Focused Attention* (FA), *Perceived Usability* (PU), *Aesthetic Appeal* (AE) and *Reward* (RW). First, the items in the PU subscale are reverse coded. The engagement score is computed by combining these twelve items. To compute the engagement score, the sum of all items is divided by twelve. The items used are:

- FA-S.1: I lost myself in this experience.
- FA-S.2: The time I spent using this search engine just slipped away.
- FA-S.3: I was absorbed in this experience.
- PU-S.1: I felt frustrated while using this search engine.
- PU-S.2: I found this search engine confusing to use.
- PU-S.3: Using this search engine was taxing.
- AE-S.1: This search engine was attractive.
- AE-S.2: This search engine was aesthetically appealing.

	Bias			
Search result position	*Disagree*	*Agree*	*Neutral (positive or neutral query)*	*Neutral (negative query)*
1	-2	2	2	-2
2	-2	2	-2	2
3	-1	1	2	-2
4	-1	1	-2	2
5	-1	1	1	-1
6	s	s	-1	1
7	s	s	1	-1
8	s	s	-1	1
9	s	s	1	-1
10	s	s	-1	1
11	-2	2	2	-2
12	-2	2	-2	2
13	-1	1	2	-2
14	-1	1	-2	2
15	-1	1	1	-1
16	s	s	-1	1
17	s	s	1	-1
18	s	s	-1	1
19	s	s	1	-1
20	s	s	-1	1

Table 5.2: Viewpoints for the search results in the search result rankings with different bias. The search results with an ’s’ as viewpoint can have any viewpoint and are based on the cosine similarity with the query. This is explained in more detail in Section 3.3.

	Condition		
Opinion towards topic	*Mitigate*	*Reinforce*	*Neutral*
Agree	Biased to disagree	Biased to agree	Neutral (variation depending on query)
Disagree	Biased to agree	Biased to disagree	Neutral (variation depending on query)

Table 5.3: Search result ranking that is shown to the participant per opinion towards the topic and condition that is assigned. For the neutral search result rankings, two different variations exist. One is chosen when the query at the top of the SERP is positive or neutral, the other one is chosen when the query is negative (see Table 5.2).

- AE-S.3: This search engine appealed to my senses.
- RW-S.1: Using this search engine was worthwhile.
- RW-S.2: My experience was rewarding.
- RW-S.3: I felt interested in this experience.

5.2.2. User engagement metrics

In Section 2.3, different metrics for UE have been described. Since the users are not allowed to issue their own queries, some of the metrics have been adjusted. The metrics that will be used here are:

- **# of Clicks**: number of SERP clicks executed during the search task.
- **# of Documents Clicked**: number of documents clicked during the search task.
- **Deepest Document Rank**: rank of the lowest document clicked.
- **Page Dwell Time**: average time spent viewing documents.
- **SATCount**: number of document clicks followed by a dwell time of at least 30 seconds.
- **SATRate**: *SATCount* divided by the number of document clicks.
- **(Active) Task Time Spent**: (active) time spent on the search task.
- **Mouseover**: number of mouseover events.
- **Pagination Frequency**: number of times the results page is changed.
- **Time to First Click**: time until the first click anywhere on the SERP is exercised.

5.2.3. Perceived diversity scale

Knijnenburg et al. [36] have introduced several items that can be used to measure the perceived diversity of recommendations. These items were adjusted to match search results instead of recommendations. The responses were recorded using a 7-point Likert scale. The five items included are:

1. I liked the search results shown by the system.
2. The search results shown fitted my preference.
3. The search results shown were relevant.
4. The search engine showed too many bad search results.
5. Overall, the search results were *not* biased towards a particular viewpoint.

The fourth item is reverse coded before computing the perceived diversity score. The score is computed by taking the average over the five items.

5.3. Variables

Independent variables

- *Topic* (categorical; between subjects). Each participant is assigned to one of the two topics; *school uniforms* or *vegetarianism*. The analysis and selection of the topics can be found in Section 3.1. The participant always has a strong viewpoint towards the topic that is assigned to the participant.

- *Condition* (categorical; between subjects). The participant is randomly assigned to one of the three conditions. The condition determines the search results that are shown to the participant. It could be chosen to *mitigate*, *reinforce* or do nothing (*neutral*) with the viewpoint. Three different search results rankings exist for each query. One of the search result rankings is biased to agree with the topic, the other is biased to disagree with the topic, while in the final condition the results are not biased. The different rankings are shown in Table 5.2. It was chosen not to use any similarity based results for the *neutral* condition since this could still induce bias. Furthermore, in the *neutral* condition, a negative query starts with a negative result, while a positive or neutral query starts with a positive search result.

- *Query* (categorical; between subjects). A query is randomly assigned to the participant. For each topic three queries exist from which one is selected. The queries are shown in Table 5.1. For each topic, one of the queries is neutral, one is positive and the last one is negative. The *query* variable indicates whether the participants' viewpoint *agrees* or *disagrees* with the stance of the query. For example, if the participant agrees that more people should become vegetarian and the query is *"The advantages of becoming vegetarian"*, the *query* variable is set to *agree*. However, if the participant would disagree with the topic and the query would be the same as previously mentioned, the *query* variable would be set to *disagree*. If the query is neutral, the *query* variable is always equal to *neutral*.

Dependent variable

- *User engagement* (continuous). User engagement is measured in two ways. As mentioned in Section 5.2.1, the UES short-form will be used to measure UE by calculating the UES. Additionally, the ten metrics mentioned in Section 5.2.2 will be used.

Covariate

- *Perceived result diversity* (continuous). The perceived viewpoint diversity of the search results is measured using the perceived diversity scale, as mentioned in Section 5.2.3.

Descriptive and exploratory measurements

- *Attitude change* (continuous). The attitude of the participants towards the topic that is assigned. The attitude is measured on a 7-point Likert scale. The attitude is also measured after the search task in order to measure attitude change.

- *Topical interest* (continuous). The interest in the topic is measured using a 7-point Likert scale. The participants had to respond to the item "I was motivated to search and learn about this topic".

- *Gender*. Participants could select their self-identified gender.

- *Age*. Participants could specify their age (in years).

5.4. Procedure

The online survey platform *Qualtrics*[1] is used to conduct the online study. The study had been approved by the Human Research Committee of TU Delft before running the study. A visual representation of the procedure of the final study after the participant has given consent and checked their Prolific ID can be found in Figure 5.2. The final study can be broken up into the following steps:

Step 1. The participants receive an introduction to the search task and indicate whether they give consent to process their data and would like to participate.

Step 2. The participants are asked to confirm their Prolific ID. The Prolific ID is filled by default with the URL parameter that is provided by Prolific.

[1] https://www.qualtrics.com/

Step 3. The participants give their opinion on the two debated topics that were selected. These are, *"More people should become vegetarian."* and *"Students should have to wear school uniforms."*. A topic is assigned to the participant. The participant always has a strong opinion on the topic that is assigned. The opinion is strong when it falls into the following set of responses: "Strongly disagree", "Disagree", "Agree" and "Strongly agree". If the participant does have a strong opinion for at least one of the topics, a condition and query will be assigned to the participant. The participant does not know which condition has been assigned. The participants will see which query has been assigned in the search interface, but does not know that other queries exist. When the participant does not have a strong opinion for any of the topics, the participant will continue to *Step 10.* In this step an attention check item is included. If the participant fails to answer the attention check item correctly, he/she will be ejected from the study.

Step 4. The following text is presented to the participants:

"Imagine the following scenario. Your government needs to decide on future policies for several debated topics. Since these topics are highly debated, they ask the people about their informed opinion. You have been selected to give your informed opinion regarding <topic>. In order to give your informed opinion you have decided to use the search engine that you will be provided with in the following question."

In this text, <topic> will be replaced with the topic that is assigned to the participant in the previous step.

Step 5. Instructions on how to correctly use the search engine are given:

"For this question, you are required to interact with the search engine for at least two minutes. There will be a bonus reward tied to the amount of time you spend using the search system. When you are done interacting with the search engine and you are able to continue to the next question, please close the search engine and then continue to the next question. If you are not allowed to proceed to the next question, please return to the search engine."

The participants are told that if they do not follow these instructions, their payment can be cancelled. Below the instructions the link to the search engine is shown. Two minutes after the link to the search engine is clicked the participant will be able to continue to the next question.

Step 6. The participants are asked to indicate how they think about each of the perceived diversity scale items.

Step 7. The participants are asked to indicate how they think about each of the UES items. Another attention check item is included between the UES items. If the participant fails to answer this item correctly, he/she will be ejected from the study.

Step 8. For exploratory purposes the participants are once again asked for their opinion towards the two topics.

Step 9. The participants indicate whether they were motivated to learn about the topic that was assigned to them.

Step 10. The participants specify their self-identified gender and age (in years).

5.5. Statistical Analyses

An *Analysis of Covariance* (ANCOVA) is performed with user engagement as dependent variable. The ANCOVA analysis has been chosen although it is not known upfront whether the data is normally distributed. It has been shown that ANOVA and ANCOVA are robust for ordinal Likert-type data that is not normally distributed [49]. *Condition*, *topic* and *query* are used as between-subjects factors. *Perceived result diversity* is used as a covariate. The hypotheses that are tested are mentioned in Section 5.1. Because four hypotheses are tested, a Bonferroni correction is applied. The new significance threshold is equal to $\alpha = \frac{0.05}{4} = 0.0125$.

5.6. Participants

The participants have been recruited via Prolific. Prescreening is used to filter out participants that are not suitable for the study. First of all, it is required that English is a fluent language of the participant. Furthermore, participants that already participated in the pilot study are not allowed to participate again in the final study. Quality control measures are applied. The two attention check items mentioned in Section 5.4 are used to eject participants from the study. Another measure that is taken to ensure that the data is of high quality is that participants are rejected when they did not actively spend enough time within the search engine. The instructions tell the participant to spend at least two minutes using the search engine. If the participant did not spend at least 90 seconds actively using the search engine, his/her submission will be rejected.

The required sample size is computed using a power analysis for the ANCOVA test. The power analysis is performed using G*Power [19]. The effect size is set to 0.25 (moderate effect). The statistical significance is set to $\alpha = 0.0125$, after applying the Bonferroni correction. The power (1 - β) is set equal to 0.8. Using these parameters the total sample size is equal to 265. These are participants that have a strong opinion for any of the two topics. It is not known how many participants have a strong opinion upfront. Therefore, the exact number of required participants is not known before the study is run.

In total, 435 participants were recruited via *Prolific*[2]. Of these 435 participants, 59 participants were rejected. In all, 41 participants incorrectly answered at least one of the two attention checks. The other 18 participants were rejected because they did not spend more than 90 seconds actively searching using the search engine. Thus, a total of 376 submissions were accepted. The average age of the participants was equal to 26, with a standard deviation of 7. The gender distribution was as follows: 51% male, 47% female and 2% other.

[2] https://www.prolific.co/

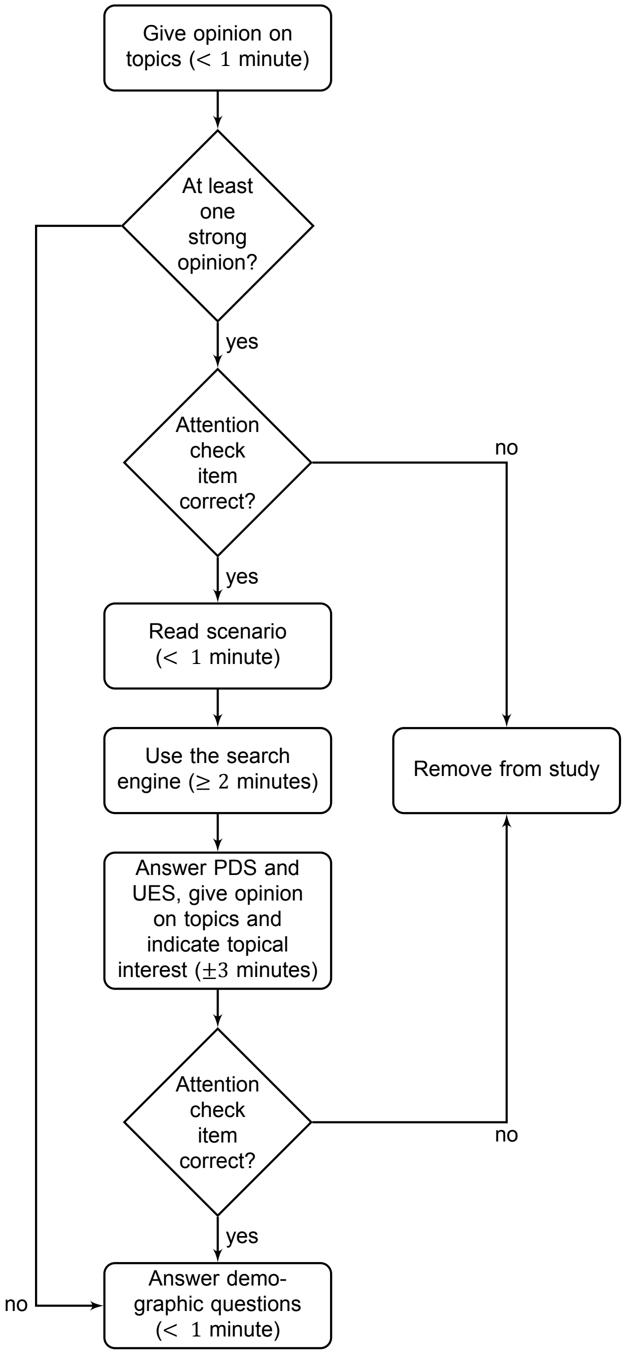

Give opinion on topics (< 1 minute)
At least one strong opinion?
yes
no
Attention check item correct?
no
yes
Read scenario (< 1 minute)
Use the search engine (≥ 2 minutes)
Remove from study
Answer PDS and UES, give opinion on topics and indicate topical interest (±3 minutes)
Attention check item correct?
no
yes
Answer demographic questions (< 1 minute)

Figure 5.2: Graphical representation of the procedure that the participants of the study follow. Each task is followed by a time estimation.

6 Results & Analysis

6.1. Descriptive Statistics

A total of 265 of the 376 participants had a strong viewpoint towards at least one of the topics. The participants were divided over the topics as follows: 134 participants were assigned the topic about *school uniforms* and 131 participants were assigned to the topic about *vegetarianism*. The participants were divided into 18 different groups. The number of participants was balanced over the different groups. For each group the sample size is shown in Table 6.1. It was attempted to balance the groups as much as possible. However, this is not always possible. For example the topic that is assigned to the user cannot always be determined. Sometimes the participant does not have a strong opinion for the topic that is underrepresented in the sample size. This is why three participants more have been recruited in the *school uniforms* topic than for the *vegetarianism* topic.

Condition	**Topic**	**Query**	*N*
Mitigate	*School Uniforms*	*Disagree*	15
		Neutral	15
		Agree	14
	Vegetarianism	*Disagree*	15
		Neutral	14
		Agree	14
Control	*School Uniforms*	*Disagree*	15
		Neutral	15
		Agree	14
	Vegetarianism	*Disagree*	15
		Neutral	15
		Agree	14
Reinforce	*School Uniforms*	*Disagree*	15
		Neutral	16
		Agree	15
	Vegetarianism	*Disagree*	15
		Neutral	14
		Agree	15

Table 6.1: Sample size for each group, where possible the group sizes were balanced.

The participants were motivated to search about their topic. The participants were told to actively interact with the search engine for at least two minutes. The actual threshold for accepting/rejecting a submission was set at 90 seconds. On average, the participants spent 4.67 minutes in the search engine with a standard deviation of 2.99 minutes. This is confirmed by the number of participants that were interested to search about the topic that was assigned to them. Of all participants, 99.62% at least somewhat agreed that they were motivated to search and learn about their assigned topic. The participants on average clicked 8.29 of the 20 search results with a standard deviation of 4.94. The majority of the participants (79.25%) requested the next page of search results.

6.2. Hypothesis Tests

UE is measured in two different ways, by *UES* and *UE metrics*. Since the variables both measure UE, the variables are not combined but tested separately. The original plan was to use the ANCOVA with *perceived result diversity* as a covariate. In order to perform ANCOVA the assumption is made that there is no interaction between the covariant and the independent variable. This is checked using an ANOVA with *condition* as independent variable and *perceived result diversity* as dependent variable. The results indicate that there is significant differences in interactions between *condition* and *perceived result diversity* ($p = 0.00001$). This can be explained by the fact that participants perceived the result diversity differently over the different conditions. It was decided to conduct ANOVAs for hypotheses **H1**-**H3**. **H4** is tested using a Multiple Regression Analysis (MRA), using *condition* and *perceived result diversity* as independent variables and *UES* as dependent variable.

6.2.1. User Engagement Score

The results of the ANOVAs are shown in Table 6.2. The ANOVA for **H1** revealed no significant direct effects of *condition* on *UES* ($F = 1.933$, $p = 0.147$, $\eta^2 = 1.540$). The tests of the interaction effects between *condition* and *topic* ($F = 0.868$, $p = 0.421$, $\eta^2 = 0.692$), and *condition* and *query* ($F = 1.190$, $p = 0.316$, $\eta^2 = 0.949$) were also not significant. An MRA was performed to test **H4**. *Condition* and *perceived result diversity* significantly predicted *UES* ($F(2, 262) = 49.036$, $p < 0.05$). *Condition* did not add statistically significantly to the prediction of *UES*. Therefore, it cannot be concluded that *perceived result diversity* moderates the effect of *condition* on *UES*.

Hypothesis	**Variables**	F	p	η^2
H1	*condition*	1.933	0.147	1.540
H2	*condition:topic*	0.868	0.421	0.692
H3	*condition:query*	1.190	0.316	0.949

Table 6.2: Results for the ANOVAs (H1-H3) with *UES* as dependent variable. Interaction effects are represented using colons.

The means over the different conditions are shown in Figure 6.1. Although the results for **H1** are not significant, a small change in the mean can be seen for the different conditions. The mean *UES* over different conditions and topics is shown in Figure 6.2. For the different conditions and queries, the *UES* is shown in Figure 6.3.

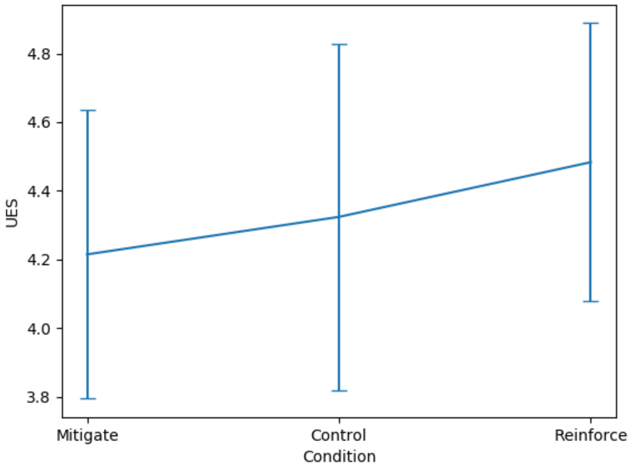


Figure 6.1: Mean *UES* for different conditions. The vertical bars represent one standard deviation. The expected trend is observed, but the differences are non-significant.

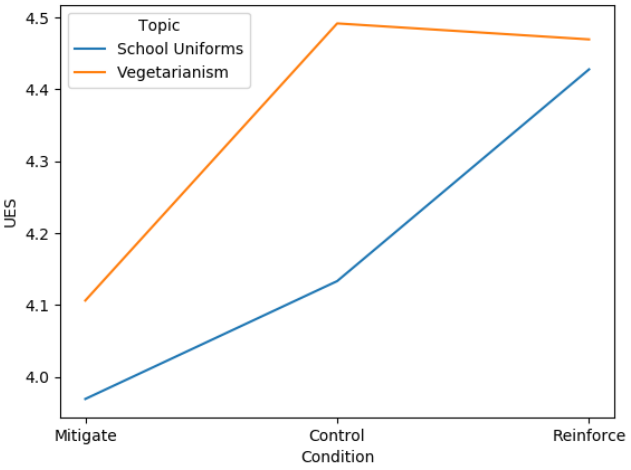


Figure 6.2: Mean *UES* for different conditions and topics.

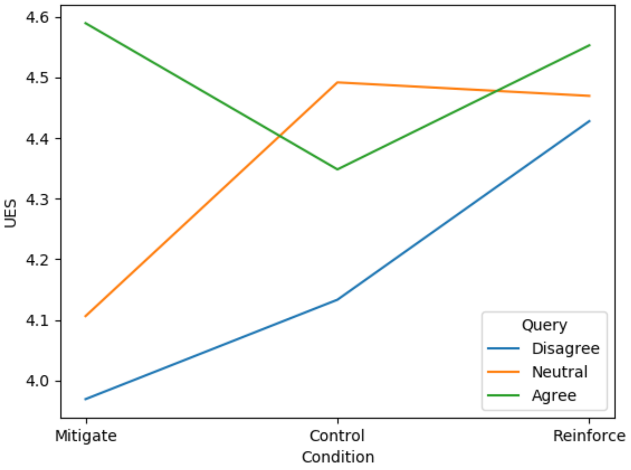


Figure 6.3: Mean *UES* for different conditions and queries. The *UES* is higher than expected where the query *agrees* with the viewpoint of the user in the *mitigate* condition.

6.2.2. User Engagement Metrics

The means and standard deviations for each UE metric per *condition* are shown in Table 6.3. For the conditions and queries these measures are shown in Table 6.4 and for the conditions and topics they are shown in Table 6.5.

Not all UE metrics are used to test for significance. The UE metrics are grouped based on the correlation between the UE metrics and are then combined. UE metrics are put into the same group when the correlation coefficient between the metric and at least one other metric in the group is at least 0.6, also known as a strong correlation [18]. The matrix with correlation coefficients is shown in Table 6.6. This means the following groups of UE metrics are created:

Group 1: *active time*, *SAT count*, *SAT rate* and *dwell time*.

Group 2: *nr. of documents clicked*, *number of mouseovers* and *pagination frequency*.

This means *time to first click*, *nr. of clicks* and *deepest document rank* are not included in the analysis. The excluded metrics are not important to the search engine providers w.r.t. increasing advertisement revenue and therefore they are not included in a separate group. For each group the

	Condition		
UE metric	*Mitigate*	*Control*	*Reinforce*
Active time (s)	***284.09*** (205.54)	***301.29*** (195.29)	***255.41*** (123.88)
Time to first click (s)	10.08 (11.51)	13.49 (10.48)	11.15 (8.78)
Nr. of clicks	30.68 (43.49)	24.77 (23.41)	28.38 (22.69)
Nr. documents clicked	9.01 (4.8)	7.09 (4.14)	8.76 (5.53)
Dwell time (s)	30.36 (25.95)	***39.87*** (32.39)	31.98 (34.51)
SAT count	2.48 (2.46)	2.6 (2.4)	2.04 (1.77)
SAT rate	0.34 (0.28)	0.43 (0.33)	0.33 (0.31)
Deepest document rank	8.31 (2.28)	7.1 (3.07)	7.33 (2.64)
Pag. freq.	1.3 (1.04)	1.15 (1.28)	1.57 (1.11)
Mouseover	55.54 (30.87)	47.47 (30.11)	58.62 (45.65)

Table 6.3: Means and standard deviations (between brackets) per metric for different conditions. It was expected that the active time in the *reinforce* condition would on average be greater than in the *mitigate* and *control* condition, but this has not been observed in the results. It is observed that the average dwell time is greater in the *control* condition.

	Condition								
	Mitigate			*Control*			*Reinforce*		
	Query			**Query**			**Query**		
UE metric	*Disagree*	*Neutral*	*Agree*	*Disagree*	*Neutral*	*Agree*	*Disagree*	*Neutral*	*Agree*
Active time (s)	274.56 (131.24)	255.53 (221.49)	323.87 (243.89)	313.69 (173.51)	294.82 (139.55)	294.94 (257.86)	268.55 (141.24)	247.87 (101.84)	249.82 (124.33)
Time to first click (s)	7.72 (8.16)	12.81 (16.13)	9.76 (7.57)	13.14 (10.0)	12.57 (10.45)	14.84 (10.87)	13.38 (10.35)	8.98 (7.95)	11.08 (7.18)
Nr. of clicks	31.5 (31.9)	21.45 (10.47)	39.36 (67.14)	27.6 (33.25)	26.7 (17.5)	19.68 (13.08)	28.83 (14.48)	28.9 (23.89)	27.4 (27.63)
Nr. documents clicked	8.73 (4.23)	8.14 (4.49)	10.21 (5.41)	6.33 (3.47)	8.03 (4.79)	6.89 (3.87)	9.3 (4.44)	9.2 (6.91)	7.77 (4.77)
Dwell time (s)	31.13 (20.94)	26.7 (16.4)	33.32 (36.33)	47.38 (39.33)	36.67 (30.61)	35.25 (23.29)	34.23 (45.7)	28.72 (23.03)	32.99 (30.62)
SAT count	2.43 (1.93)	2.28 (2.9)	2.75 (2.46)	2.43 (1.76)	2.83 (2.15)	2.54 (3.11)	1.87 (1.69)	2.23 (1.84)	2.03 (1.76)
SAT rate	0.38 (0.3)	0.33 (0.26)	0.32 (0.27)	0.47 (0.35)	0.42 (0.33)	0.39 (0.28)	0.29 (0.3)	0.35 (0.32)	0.36 (0.3)
Deepest document rank	7.93 (1.44)	9.0 (3.28)	8.0 (1.44)	6.87 (3.61)	7.07 (2.29)	7.39 (3.14)	7.77 (1.76)	7.1 (2.43)	7.13 (3.41)
Pag. freq.	1.17 (1.04)	1.38 (1.06)	1.36 (1.01)	1.4 (1.72)	0.8 (0.83)	1.25 (1.02)	1.83 (1.13)	1.6 (1.2)	1.27 (0.89)
Mouseover	48.73 (32.89)	55.21 (28.0)	63.18 (29.69)	47.33 (35.22)	47.67 (29.83)	47.39 (23.8)	64.57 (37.98)	62.6 (61.69)	48.7 (29.23)

Table 6.4: Mean and standard deviations (between brackets) per metric for different conditions and queries. No consistent differences have been observed over the different conditions and queries.

metrics are then combined by taking the average. It would have been better to combine the metrics using different weights since it is unlikely that all metrics contribute the same to UE. However, these weights are not known and also cannot be fit on another measure.

The hypothesis tests from the previous section are repeated with the combined UE metrics as dependent variables. The mean and standard deviation of the first combined metric are shown in Figure 6.4. The results of the ANOVA analyses are shown in Table 6.7. The main effect of *condition* ($F = 2.566$, $p = 0.079$, $\eta^2 = 0.055$), the interaction effect between *condition* and *topic* ($F = 3.924$,

	Condition					
	Mitigate		*Control*		*Reinforce*	
	Topic		**Topic**		**Topic**	
UE metric	*School uniforms*	*Vegetarianism*	*School uniforms*	*Vegetarianism*	*School uniforms*	*Vegetarianism*
Active time (s)	276.17 (192.45)	292.19 (217.83)	***339.43*** (239.27)	263.15 (126.95)	220.79 (82.76)	291.61 (147.19)
Time to first click (s)	10.32 (13.14)	9.82 (9.56)	13.69 (9.94)	13.28 (10.99)	10.81 (7.94)	11.49 (9.57)
Nr. of clicks	28.25 (22.54)	33.16 (57.39)	27.32 (30.43)	22.23 (12.51)	29.54 (27.11)	27.16 (16.8)
Nr. documents clicked	10.02 (5.25)	7.98 (4.05)	6.5 (3.8)	7.68 (4.38)	9.0 (6.4)	8.5 (4.42)
Dwell time (s)	29.17 (29.92)	31.57 (21.06)	***45.71*** (33.65)	34.03 (29.96)	29.32 (27.44)	34.77 (40.42)
SAT count	2.2 (1.95)	2.77 (2.87)	3.05 (2.66)	2.16 (2.0)	1.67 (1.38)	2.43 (2.03)
SAT rate	0.31 (0.27)	0.38 (0.29)	0.5 (0.31)	0.36 (0.32)	0.32 (0.31)	0.34 (0.3)
Deepest document rank	8.2 (0.99)	8.42 (3.08)	6.66 (3.16)	7.55 (2.91)	7.02 (2.42)	7.66 (2.82)
Pag. freq.	1.3 (1.01)	1.3 (1.07)	1.14 (1.41)	1.16 (1.15)	1.35 (1.07)	1.8 (1.1)
Mouseover	56.64 (34.05)	54.42 (27.18)	48.2 (31.6)	46.73 (28.52)	58.33 (48.09)	58.93 (42.95)

Table 6.5: Means and standard deviations (between brackets) per metric for different conditions and topics. It is observed that the active time in the *control* condition for the *school uniforms* topic is greater than the other active times. Just as for the *control* condition in Table 6.3, it is observed that the dwell time for the *control* condition and *school uniforms* topic is greater than for the other conditions and topics.

	Active time	**Time to first click**	**Nr. of clicks**	**Nr. documents clicked**	**Dwell time**	**SAT count**	**SAT rate**	**Deepest document rank**	**Pag. freq.**	**Mouseover**
Active time	1.0	0.07	0.23	0.27	0.43	0.79	0.41	0.06	0.09	0.22
Time to first click	0.07	1.0	-0.09	-0.03	0.03	0.01	0.05	0.06	0.07	0.06
Nr. of clicks	0.23	-0.09	1.0	0.39	-0.04	0.19	-0.04	0.12	0.21	0.38
Nr. documents clicked	0.27	-0.03	0.39	1.0	-0.36	0.2	-0.38	0.35	0.31	0.67
Dwell time	0.43	0.03	-0.04	-0.36	1.0	0.38	0.8	-0.25	-0.17	-0.24
SAT count	0.79	0.01	0.19	0.2	0.38	1.0	0.6	0.04	-0.03	0.12
SAT rate	0.41	0.05	-0.04	-0.38	0.8	0.6	1.0	-0.25	-0.24	-0.28
Deepest document rank	0.06	0.06	0.12	0.35	-0.25	0.04	-0.25	1.0	0.12	0.22
Pag. freq.	0.09	0.07	0.21	0.31	-0.17	-0.03	-0.24	0.12	1.0	0.64
Mouseover	0.22	0.06	0.38	0.67	-0.24	0.12	-0.28	0.22	0.64	1.0

Table 6.6: Pearson's r between the different UE metrics. Metrics with a strong correlation ($r \geq 0.6$) are grouped for the analyses.

$p = 0.021$, $\eta^2 = 0.083$), and the interaction effect between *condition* and *query* ($F = 0.304$, $p = 0.875$, $\eta^2 = 0.006$) on the combined UE metric of the first group are all non-significant. The multiple regression analysis shows that *condition* and *perceived result diversity* do not statistically significantly predict the combined UE metric of the first group ($F(2, 262) = 1.878, p = 0.155$).

Hypothesis	**Variables**	F	p	η^2
H1	*condition*	2.566	0.079	0.055
H2	*condition:topic*	3.924	0.021	0.083
H3	*condition:query*	0.304	0.875	0.006

Table 6.7: Results for the ANOVA analyses with the combination of the first group of UE metrics as dependent variable. Interaction effects are represented using colons.

The mean and standard deviation of the combined metric of the second group are shown in Figure 6.5. The results of the ANOVAs for the combined metric of the second group are shown in Table 6.8. The main effect of *condition* ($F = 3.730$, $p = 0.025$, $\eta^2 = 0.058$), the interaction effect between *condition* and *topic* ($F = 0.573$, $p = 0.565$, $\eta^2 = 0.009$), and the interaction effect between *condition* and *query* ($F = 1.453$, $p = 0.217$, $\eta^2 = 0.022$) on the combined UE metric of the second group are

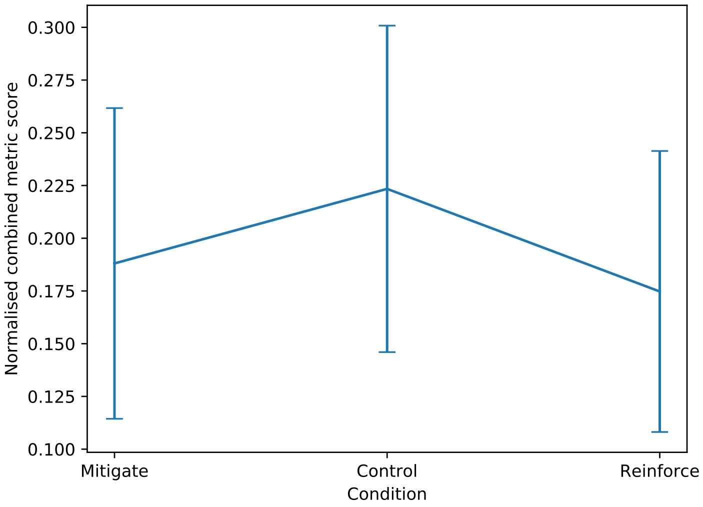


Figure 6.4: Mean combined UE metric of the first group for different conditions and queries. The vertical bars represent one standard deviation.

all non-significant. The MRA shows that *condition* and *perceived result diversity* do not statistically significantly predict the combined UE metric of the second group ($F(2, 262) = 1.040, p = 0.355$).

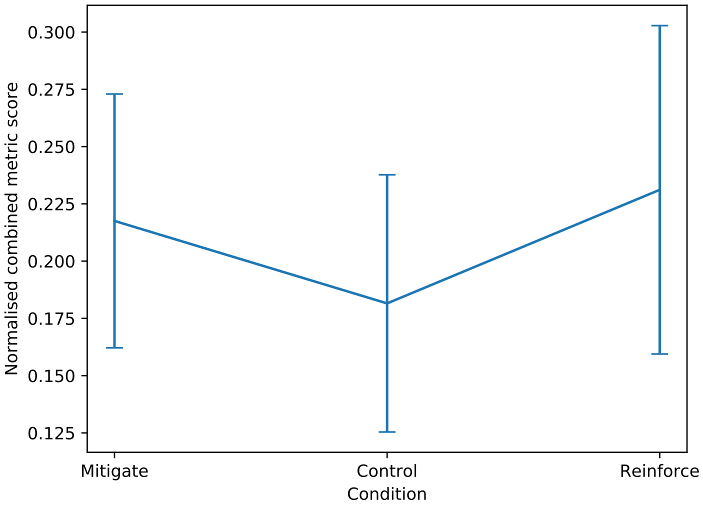


Figure 6.5: Mean combined UE metric of the second group for different conditions and queries. The vertical bars represent one standard deviation.

Hypothesis	**Variables**	F	p	η^2
H1	*condition*	3.730	0.025	0.058
H2	*condition:topic*	0.573	0.565	0.009
H3	*condition:query*	1.453	0.217	0.022

Table 6.8: Results for the ANOVAs with the combination of the second group of UE metrics as dependent variable. Interaction effects are represented using colons.

6.3. Exploratory Findings

The results show that there is no significant effect of *condition* on *UES*. When looking at Figure 6.1, the expected upward trend over the different conditions can be seen, however. In this section exploratory analyses will be performed to understand how the most interesting results came about.

6.3.1. Viewpoint-Aligned Proportion

When looking at the results shown in Figure 6.3, one of the most interesting results is the high average *UES* in the *mitigate* condition when the stance of the query is in line with the viewpoint of the user. One potential explanation of this interesting result is that participants only interacted with the search results they liked. In order to investigate whether this is the cause of the interesting result, the proportion of clicks on viewpoint-aligning documents among all document clicks will be analysed. If the proportion viewpoint aligning clicks is significantly higher in the *mitigate* condition for the *agree* query condition than for the other conditions, this could explain the results achieved. The proportion of viewpoint-aligning documents among all document clicks are shown in Table 6.9. In order to check whether the difference in proportion of viewpoint-aligned results among all clicked results between different queries is significant, a Two-Way ANOVA is performed with *query* and *condition* as independent variables. The Two-Way ANOVA showed no significant change in proportion of viewpoint-aligned documents clicked for the interaction between *condition* and *query* ($F = 1.468$, $p = 0.212$, $\eta^2 = 0.041$). The results show that the effect of *condition* on the proportion of viewpoint-aligned documents among all documents clicked is statistically significant ($F = 62.123$, $p = 0.000$, $\eta^2 = 1.750$).

	Query		
Condition	*Disagree*	*Neutral*	*Agree*
Mitigate	0.36 (0.13)	0.40 (0.15)	0.41 (0.12)
Control	0.51 (0.15)	0.55 (0.16)	0.54 (0.21)
Reinforce	0.61 (0.16)	0.62 (0.20)	0.79 (0.13)

Table 6.9: Proportion of viewpoint-aligned documents among all clicked documents over different conditions and queries.

6.3.2. Effect of Query

Because of the interesting behaviour of *UES* for the queries over the different conditions as seen in Figure 6.3, it is worth looking at the effect of the *query* on *UES*. The mean *UES* for the different queries across all conditions is shown in Figure 6.6. The effect of *query* on *UES* is tested using a One-Way ANOVA. The analysis shows no significant effect of *query* on *UES* ($F = 2.864$, $p = 0.059$, $\eta^2 = 2.287$).

6.3.3. Difference in Dwell Time

When looking at Table 6.3, a difference in *dwell time* can be seen between the *biased* conditions and the *control* condition. In order to test whether this difference is statistically significant a One-Way ANOVA is performed. Here, the biased conditions are grouped based on whether the SERP is biased (*mitigate* and *reinforce*) or not biased (*control*), this is used as an independent variable. The average *dwell time* is used as the dependent variable. The results of the analysis indicate that mean *dwell time* is statistically significantly different at a significance level of $\alpha = 0.05$ depending on whether the SERP is biased ($F = 4.517$, $p = 0.035$, $\eta^2 = 4434.299$). It is also tested what type of documents are being clicked. The proportion of high-quality documents (with viewpoint ranking -2 or 2) among all clicked documents is used as a dependent variable. The biased conditions are grouped again and this new variable is used as an independent variable. The One-Way ANOVA shows there is a significant difference between different groups in the proportion of high-quality documents clicked ($F = 87.711$, $p = 0.000$, $\eta^2 = 3.329$). The mean for the control group is equal to 0.60, while the mean for the biased group is equal to 0.36.

6.3.4. Ranks Clicked

Figure 6.7 shows for each rank which proportion of the participants clicked that rank at least once over the different conditions. The ranks of the search results are the ranks within the SERP. So, if the first result on the second page is clicked, this is considered as a click on the first rank. The ranks that are

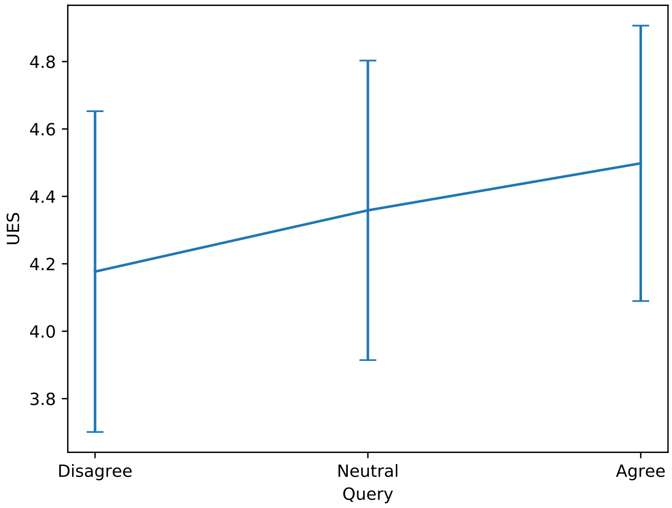


Figure 6.6: Mean *UES* for different queries (across all conditions). The vertical bars represent one standard deviation.

clicked are similar over the conditions for the first ranks. For the ranks lower on the page the proportion of participants that clicked is higher in the *biased* conditions than in the *control* condition.

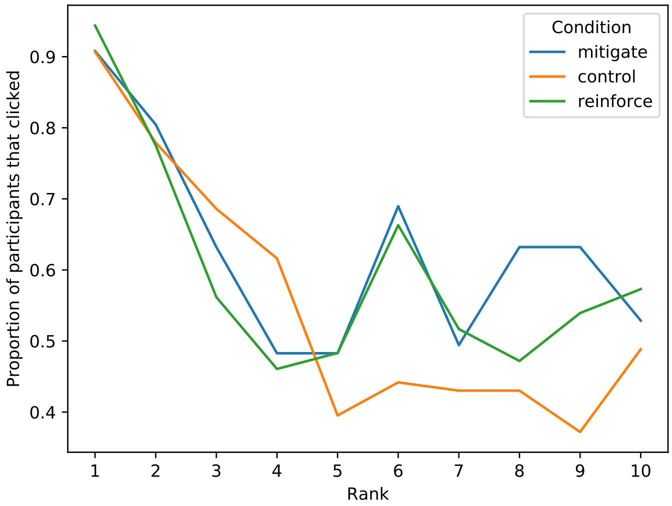


Figure 6.7: Proportion of participants that clicked on a search result rank.

6.3.5. Second Combined UE Metric

The combined UE metric for the second group of UE metrics seems to be greater in the biased conditions than in the *control* condition (Figure 6.5). A One-Way ANOVA between the biased conditions as one group and the *control* condition with the combined UE metric for the second group as dependent variable is performed. The results show that there is a significant difference in the combined UE metric for the second group between the biased conditions and the control *condition* ($F = 7.053$, $p = 0.008$, $\eta^2 = 0.108$).

7 Discussion

In this chapter the results shown in the previous chapter will be discussed. First, the key findings will be discussed in Section 7.1. In Section 7.2, the results regarding cognitive biases will be discussed. Thirdly, the difference in search result interactions between the biased conditions and the *control* condition will be discussed in Section 7.4. Then, the difference between measures of UE will be discussed in Section 7.3. Finally, in Section 7.5, limitations, weaknesses and future work will be discussed.

7.1. Key Findings

It was expected that that introducing bias in the SERP would affect the engagement of users with the search engine. It was thought that reinforcing bias in the SERP would lead to an increase in UE compared to the *control* condition, while mitigating bias in the SERP would lead to a decrease in UE compared to the *control* condition. However, the results suggest that there is no significant difference measured in UE over the different conditions, both for *UES* and *combined UE metrics*. This indicates that UE is not affected by the viewpoints that are shown to user first. Instead, the results have shown that the clicks on the results seem to be guided by the position of the search results clicked (Figure 6.7 on page 38).

7.2. Cognitive Biases

It was expected that participants would be less engaged when bias is mitigated and more engaged when bias is reinforced due to *cognitive dissonance*. The term *cognitive dissonance* describes the mental discomfort that people experience when their beliefs are challenged [12, 21, 27]. As a result of *cognitive dissonance* people tend to avoid conflicting beliefs. *Confirmation bias* can be seen as a concept that is related to *cognitive dissonance*. *Confirmation bias* causes users to dismiss or disregard information when this information is not in line with their own beliefs [35, 65]. In the exploratory analysis it was found that the proportion of viewpoint aligning results among all clicked results was significantly different over the conditions. This result indicates that confirmation bias did not occur, at least not to a high degree. This result is line with the results found by Draws et al. [13]. If *confirmation bias* would occur, the expectation would be that the difference in the proportion of viewpoint-aligning documents clicked over the conditions would be non-significant. It would also be expected that the proportion of viewpoint-aligned documents among all clicked documents would deviate from the proportion of viewpoint-aligned documents among all documents displayed in the SERP. In the *mitigate* condition the proportion of displayed viewpoint-aligning documents was on average equal to 0.4, in the *control* condition 0.5 and in the *reinforce* condition 0.6. No alarming differences are found when comparing these values with the values shown in Table 6.9 on page 37.

It was furthermore expected that users would actively spend the most time when their bias was reinforced. However, it was found that the *active task time spent* was the highest in the *control* condition. The exploratory analysis showed that there is a statistically significant difference in *dwell time* between the conditions with bias and the *control* condition. This can be explained by the significant difference in the proportion of high-quality documents clicked among all clicked documents. The high-quality documents require more effort to read. The low-quality documents are often short lists which can be quickly scanned, while the high-quality documents need to be read with more cognitive effort in order to comprehend the document. The exploratory analysis showed that the proportion of high-quality documents clicked was significantly larger in the *control* condition than in the conditions with bias. In the *control* condition more high-quality search results are shown than in the conditions with bias. 40% of the documents in the *control* condition are of high quality (four at the top of each page), while this is on average 22% in the biased conditions (two at the top of each page and possibly some in the last five results of each page). This also shows that in both the biased conditions and the *control* condition, the high-quality results are relatively clicked more often than they are present in the SERP. This indicates that *position bias* existed, although this could also be caused by the fact that users are more likely to click high-quality results. Relative to the availability, the users in the biased condition click more high-quality documents.

Low-quality documents are very similar in terms of their content. It is suspected that because more low-quality documents are shown in the biased conditions, users search for more variation which is achieved by reading higher quality documents. Some *position bias* is also confirmed by Figure 6.7 on page 38, since the ranks clicked are similar over the different conditions and the higher ranks are clicked more often. In the biased conditions a large proportion of the participants clicked on search results lower on the page. It is expected that this is caused by less high-quality results shown on the SERP in the biased conditions. It is not suspected this is due to *confirmation bias*. If this would have been caused by *confirmation bias*, this would only lead to an increase in the proportion of the participants that clicked the ranks lower in the page for the *mitigate* condition and not for the *reinforce* condition. Since for the *mitigate* condition the search results that agree with the participant appear at the bottom of the page while for the *reinforce* condition they appear at the top of the page.

Why are users not less engaged when bias is mitigated compared to when bias is reinforced? It seems that users are not affected by the viewpoints of the results that are presented in the SERP. No evidence has been found that shows a confirmation bias existed. Instead, the evidence suggests that the participants did not have a *confirmation bias*. Rather, the documents that are clicked seem to be guided by a *position bias*.

7.3. Implications for Search Engine Providers

Whereas the expected trend over the different conditions was observed for *UES* (although statistically non-significant), this was not the case for the combined UE metrics. Both measures of UE are important for search engine providers. *UES* is a subjective measure that assigns a score to how the user experiences their interaction with a system [51]. Chances are that when the user assigns a higher UES to a system, the user is more likely to use that system again in the future. The UE metrics show the behaviour of the user. This can be interesting for the search engine provider since when a user for example visits more websites, he/she creates more revenue for the search engine provider. One would expect that *UES* and *UE metrics* are correlated. For example, someone that assigns a high UES is willing to spend more time in the search engine than someone that assigns a low UES. Or, when a user likes the search results shown by the search engine more, he/she wants to view more search results. However, as mentioned before, the same trend has not been observed over the conditions for UES and the combined UE metrics score. This could be due to the composition of the SERP in terms of the quality of the search results.

7.4. More Search Result Interactions in Biased Conditions

In Section 6.3.5 it was found that the combined UE metric for the second group of UE metrics was significantly greater in the biased conditions than in the *control* condition. The second group of UE metrics consists of the *nr. of documents clicked*, the *number of mouseovers* and the *pagination frequency*. If in reality introducing bias could lead to an increase for these metrics, this could be motivation for search engine providers to manipulate search results. However, it is not expected that this increase was caused by the bias in the SERP but by the fact that more low-quality search results that are similar in terms of content were shown in the SERPs. The *SAT rate* in Table 6.3 on page 34 is lower in the biased conditions than in the *control* condition. Users were thus less satisfied with the search results in the biased conditions than in the *control* condition. This likely caused users in the biased conditions to look for better search results, leading to an increase in the combined UE metric for the second group of UE metrics.

7.5. Limitations, Caveats and Weaknesses

Several limitations have been identified with this thesis. The first and the most important limitation is that users were not allowed to issue their own queries. Letting the participant issue their own queries, makes the user conscious about the search process. This could for example change the expectations of the participants about the results that will be shown. As a result, this could then also change the engagement of the user. Also, the results were presented in plain HTML such that search results had only one viewpoint and potential domain biases were avoided. This could have had an impact on how realistic the users interpreted the search experience. A weakness that was detected when analysing the results was that for some log events the result rank had not been logged. It is unclear why this happened and not expected that this caused any problems since this happened only for 31 out of the 2649 result-click log events.

8 Conclusions

The goal of this thesis was to find out how UE is affected in users with strong opinions by bias in the SERP. The overarching research question is: *How is UE affected in users with strong opinions when SERPs are biased?* It is investigated how reinforcing and mitigating bias by manipulating the SERP affects UE. SERPs are manipulated by changing the viewpoints in different positions of the SERPs. A search engine has been developed and used in a $2 \times 3 \times 3$ factorial user study to investigate the effect of bias on UE. The results show that UE in terms of both *UES* and *combined UE metrics* is not affected by manipulating the SERP. This was shown in sections 6.2.1 and 6.2.2 for *UES* and *combined UE metrics*, respectively.

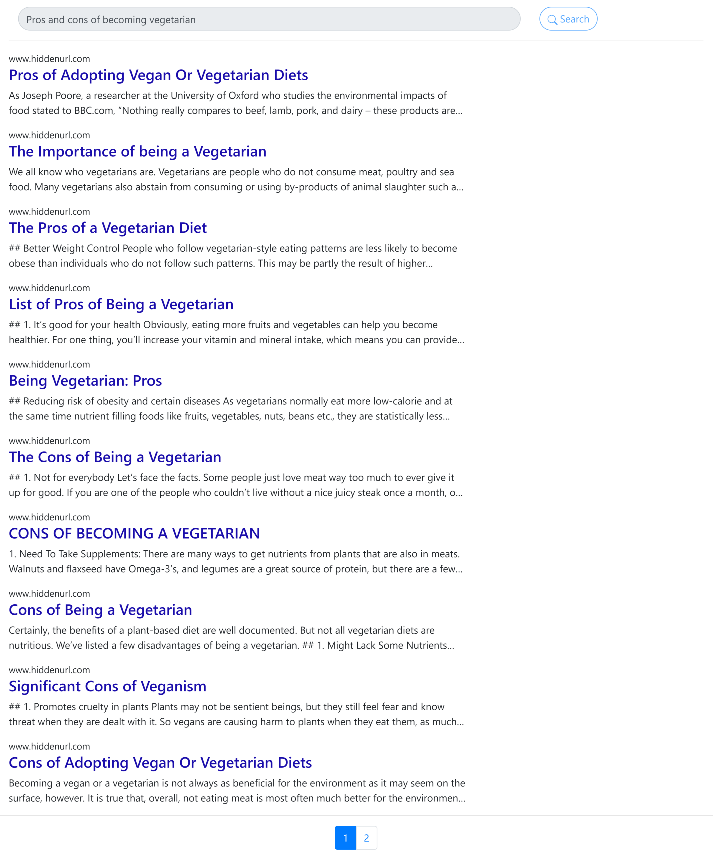
It was expected that mitigating bias would lead to a decrease in UE, while reinforcing bias would lead to an increase in UE. It is thus concluded that mitigating bias in the SERP had no impact on UE in users with strong opinions (**RQ1**). Reinforcing bias in the SERP also had no impact on UE in users with strong opinions (**RQ2**). It was thought that mitigating and/or reinforcing bias in SERPs would change *cognitive dissonance*. It was expected that users would have a *confirmation bias*, causing users to be less satisfied with the SERP when bias is mitigated and more satisfied with the SERP when bias is reinforced. However, as discussed in Section 7.2, no signs of *confirmation bias* were found. Also discussed in this section is that findings suggest that the clicks on search results are guided by *position bias* instead of *confirmation bias*, as is shown in Figure 6.7 on page 38.

By answering both research questions it has been shown that UE is not affected when SERPs are biased for the debated topics of *school uniforms* and *vegetarianism*. This shows that there is no incentive for search engine providers regarding UE to manipulate search results for these debated topics.

8.1. Future Work

Based on the limitations of this thesis there are several recommendations for future work. One of the limitations of this thesis is the aggregation of the UE metrics. In this thesis the UE metrics are aggregated by taking the average of the metrics that have been selected (i.e., each selected metric received equal weight). In reality it is not likely that each metric has the same weight, therefore it is recommended to investigate how these metrics can be best aggregated. This can be done both from the viewpoint of the search engine provider as well as the viewpoint of the user, depending on the use case of the combined UE metric. To build further upon this thesis it would be best to investigate the aggregation from the viewpoint of the search engine provider, since the goal is to find out whether there is incentive in terms of UE for the search engine provider to manipulate SERPs. Another limitation of this work was that users were not allowed to issue their own queries. The users were presented a SERP with a fixed query and the corresponding search results. In reality, users issue their own queries. The process of users either issuing their own query or not issuing their own query could affect UE. For example, if users issue their own query they might be less satisfied with biased SERPs than when the query is chosen for them. It is advised to investigate this effect by running a similar study where users have more freedom regarding their query or queries. In this new study it is important to think about how search results will be ranked. It is important to think about the amount of freedom the user will get. This has also been one of the main considerations of this thesis. One needs to think about how to rank the search results w.r.t. the query that is issued by the user such that it is still possible to analyse the results. But, the search results should also match the query in order to make the search process realistic (i.e., the search results should not be the same for all queries that the user issues). A step between this work and letting the users issue their own queries would be to let users pick their own query from a set of queries. This gives the choice of picking the query to the user while also controlling which results will be shown to the user. It is also recommended to make the search results more realistic than the search results in this thesis. This means the search results could be shown using HTML format with styling instead of plain HTML to provide more variety. Also, most low-quality search results were very similar in terms of content and this could be avoided by changing the content of these search results and/or changing the retrieval algorithm for search results. This could result in a change in UE (e.g., a change in the amount of documents clicked).

A
Figures



Pros and cons of becoming vegetarian

www.hiddenurl.com
Pros of Adopting Vegan Or Vegetarian Diets
As Joseph Poore, a researcher at the University of Oxford who studies the environmental impacts of food stated to BBC.com, "Nothing really compares to beef, lamb, pork, and dairy – these products are...

www.hiddenurl.com
The Importance of being a Vegetarian
We all know who vegetarians are. Vegetarians are people who do not consume meat, poultry and sea food. Many vegetarians also abstain from consuming or using by-products of animal slaughter such a...

www.hiddenurl.com
The Pros of a Vegetarian Diet
Better Weight Control People who follow vegetarian-style eating patterns are less likely to become obese than individuals who do not follow such patterns. This may be partly the result of higher...

www.hiddenurl.com
List of Pros of Being a Vegetarian
1. It's good for your health Obviously, eating more fruits and vegetables can help you become healthier. For one thing, you'll increase your vitamin and mineral intake, which means you can provide...

www.hiddenurl.com
Being Vegetarian: Pros
Reducing risk of obesity and certain diseases As vegetarians normally eat more low-calorie and at the same time nutrient filling foods like fruits, vegetables, nuts, beans etc., they are statistically less...

www.hiddenurl.com
The Cons of Being a Vegetarian
1. Not for everybody Let's face the facts. Some people just love meat way too much to ever give it up for good. If you are one of the people who couldn't live without a nice juicy steak once a month, o...

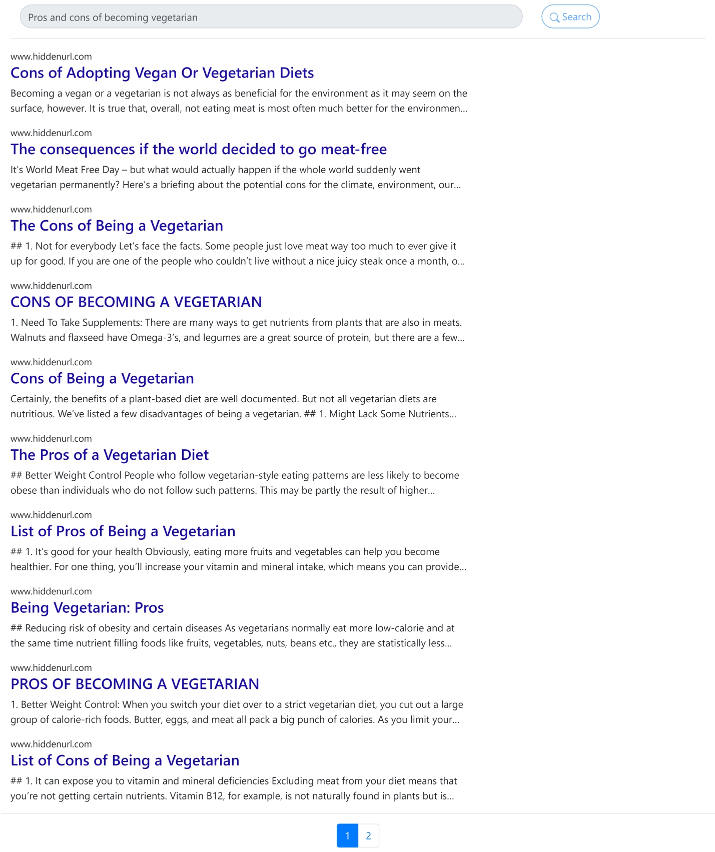
www.hiddenurl.com
CONS OF BECOMING A VEGETARIAN
1. Need To Take Supplements: There are many ways to get nutrients from plants that are also in meats. Walnuts and flaxseed have Omega-3's, and legumes are a great source of protein, but there are a few...

www.hiddenurl.com
Cons of Being a Vegetarian
Certainly, the benefits of a plant-based diet are well documented. But not all vegetarian diets are nutritious. We've listed a few disadvantages of being a vegetarian. ## 1. Might Lack Some Nutrients...

www.hiddenurl.com
Significant Cons of Veganism
1. Promotes cruelty in plants Plants may not be sentient beings, but they still feel fear and know threat when they are dealt with it. So vegans are causing harm to plants when they eat them, as much...

www.hiddenurl.com
Cons of Adopting Vegan Or Vegetarian Diets
Becoming a vegan or a vegetarian is not always as beneficial for the environment as it may seem on the surface, however. It is true that, overall, not eating meat is most often much better for the environmen...

Figure A.1: SERP with the "Pros and cons of becoming vegetarian" query for which the search results have been biased to *agree*.



Pros and cons of becoming vegetarian

www.hiddenurl.com

Cons of Adopting Vegan Or Vegetarian Diets

Becoming a vegan or a vegetarian is not always as beneficial for the environment as it may seem on the surface, however. It is true that, overall, not eating meat is most often much better for the environmen...

www.hiddenurl.com

The consequences if the world decided to go meat-free

It's World Meat Free Day – but what would actually happen if the whole world suddenly went vegetarian permanently? Here's a briefing about the potential cons for the climate, environment, our...

www.hiddenurl.com

The Cons of Being a Vegetarian

1. Not for everybody Let's face the facts. Some people just love meat way too much to ever give it up for good. If you are one of the people who couldn't live without a nice juicy steak once a month, o...

www.hiddenurl.com

CONS OF BECOMING A VEGETARIAN

1. Need To Take Supplements: There are many ways to get nutrients from plants that are also in meats. Walnuts and flaxseed have Omega-3's, and legumes are a great source of protein, but there are a few...

www.hiddenurl.com

Cons of Being a Vegetarian

Certainly, the benefits of a plant-based diet are well documented. But not all vegetarian diets are nutritious. We've listed a few disadvantages of being a vegetarian. ## 1. Might Lack Some Nutrients...

www.hiddenurl.com

The Pros of a Vegetarian Diet

Better Weight Control People who follow vegetarian-style eating patterns are less likely to become obese than individuals who do not follow such patterns. This may be partly the result of higher...

www.hiddenurl.com

List of Pros of Being a Vegetarian

1. It's good for your health Obviously, eating more fruits and vegetables can help you become healthier. For one thing, you'll increase your vitamin and mineral intake, which means you can provide...

www.hiddenurl.com

Being Vegetarian: Pros

Reducing risk of obesity and certain diseases As vegetarians normally eat more low-calorie and at the same time nutrient filling foods like fruits, vegetables, nuts, beans etc., they are statistically less...

www.hiddenurl.com

PROS OF BECOMING A VEGETARIAN

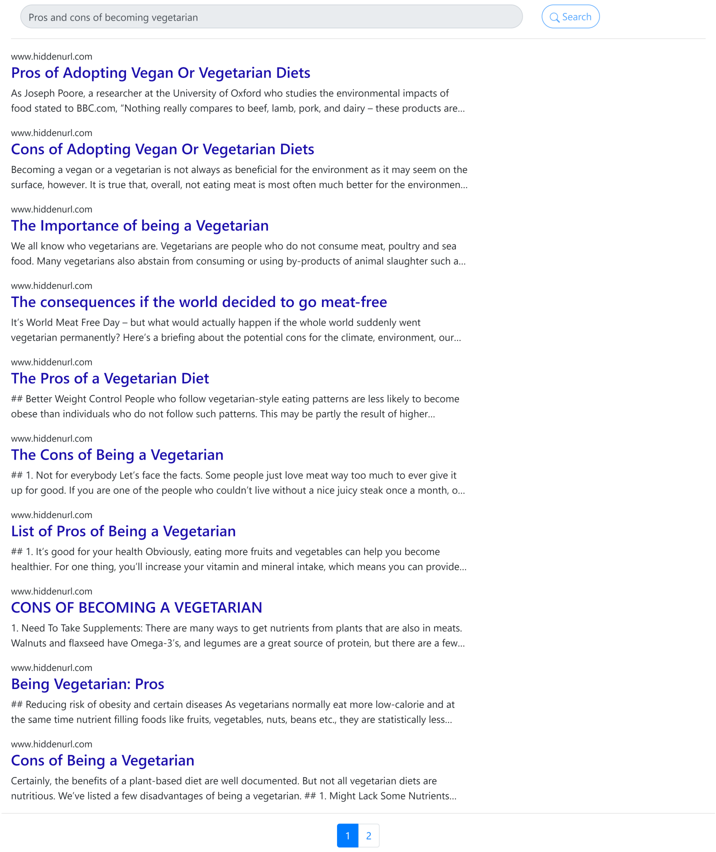
1. Better Weight Control: When you switch your diet over to a strict vegetarian diet, you cut out a large group of calorie-rich foods. Butter, eggs, and meat all pack a big punch of calories. As you limit your...

www.hiddenurl.com

List of Cons of Being a Vegetarian

1. It can expose you to vitamin and mineral deficiencies Excluding meat from your diet means that you're not getting certain nutrients. Vitamin B12, for example, is not naturally found in plants but is...

Figure A.2: SERP with the “Pros and cons of becoming vegetarian” query for which the search results have been biased to *disagree*.



Pros and cons of becoming vegetarian

www.hiddenurl.com
Pros of Adopting Vegan Or Vegetarian Diets
As Joseph Poore, a researcher at the University of Oxford who studies the environmental impacts of food stated to BBC.com, "Nothing really compares to beef, lamb, pork, and dairy – these products are...

www.hiddenurl.com
Cons of Adopting Vegan Or Vegetarian Diets
Becoming a vegan or a vegetarian is not always as beneficial for the environment as it may seem on the surface, however. It is true that, overall, not eating meat is most often much better for the environmen...

www.hiddenurl.com
The Importance of being a Vegetarian
We all know who vegetarians are. Vegetarians are people who do not consume meat, poultry and sea food. Many vegetarians also abstain from consuming or using by-products of animal slaughter such a...

www.hiddenurl.com
The consequences if the world decided to go meat-free
It's World Meat Free Day – but what would actually happen if the whole world suddenly went vegetarian permanently? Here's a briefing about the potential cons for the climate, environment, our...

www.hiddenurl.com
The Pros of a Vegetarian Diet
Better Weight Control People who follow vegetarian-style eating patterns are less likely to become obese than individuals who do not follow such patterns. This may be partly the result of higher...

www.hiddenurl.com
The Cons of Being a Vegetarian
1. Not for everybody Let's face the facts. Some people just love meat way too much to ever give it up for good. If you are one of the people who couldn't live without a nice juicy steak once a month, o...

www.hiddenurl.com
List of Pros of Being a Vegetarian
1. It's good for your health Obviously, eating more fruits and vegetables can help you become healthier. For one thing, you'll increase your vitamin and mineral intake, which means you can provide...

www.hiddenurl.com
CONS OF BECOMING A VEGETARIAN
1. Need To Take Supplements: There are many ways to get nutrients from plants that are also in meats. Walnuts and flaxseed have Omega-3's, and legumes are a great source of protein, but there are a few...

www.hiddenurl.com
Being Vegetarian: Pros
Reducing risk of obesity and certain diseases As vegetarians normally eat more low-calorie and at the same time nutrient filling foods like fruits, vegetables, nuts, beans etc., they are statistically less...

www.hiddenurl.com
Cons of Being a Vegetarian
Certainly, the benefits of a plant-based diet are well documented. But not all vegetarian diets are nutritious. We've listed a few disadvantages of being a vegetarian. ## 1. Might Lack Some Nutrients...

Figure A.3: SERP with the "Pros and cons of becoming vegetarian" query and the *neutral* search result ranking.

Bibliography

[1] Ahmed Allam, Peter Johannes Schulz, and Kent Nakamoto. 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):e100, apr 2014. ISSN 14388871. doi: 10.2196/jmir.2642. URL `https://www.jmir.org/2014/4/e100/`.

[2] Ioannis Arapakis and Luis A. Leiva. Predicting User Engagement with Direct Displays Using Mouse Cursor Information. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 599–608, New York, NY, USA, jul 2016. ACM. ISBN 9781450340694. doi: 10.1145/2911451.2911505. URL `https://dl.acm.org/doi/10.1145/2911451.2911505`.

[3] Ioannis Arapakis, Mounia Lalmas, and George Valkanas. Understanding Within-Content Engagement through Pattern Analysis of Mouse Gestures. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1439–1448, New York, NY, USA, nov 2014. ACM. ISBN 9781450325981. doi: 10.1145/2661829.2661909. URL `https://dl.acm.org/doi/10.1145/2661829.2661909`.

[4] Josh Attenberg, Sandeep Pandey, and Torsten Suel. Modeling and Predicting User Behavior in Sponsored Search. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*, page 1067, New York, New York, USA, 2009. ACM Press. ISBN 9781605584959. doi: 10.1145/1557019.1557135. URL `http://portal.acm.org/citation.cfm?doid=1557019.1557135`.

[5] Leif Azzopardi. Cognitive Biases in Search. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, pages 27–37, New York, NY, USA, mar 2021. ACM. ISBN 9781450380553. doi: 10.1145/3406522.3446023. URL `https://dl.acm.org/doi/10.1145/3406522.3446023`.

[6] Ricardo Baeza-Yates. Data and Algorithmic Bias in the Web. In *Proceedings of the 8th ACM Conference on Web Science - WebSci '16*, pages 1–1, New York, New York, USA, 2016. Association for Computing Machinery (ACM). doi: 10.1145/2908131.2908135. URL `http://dl.acm.org/citation.cfm?doid=2908131.2908135`.

[7] Nick Bansback, Linda C. Li, Larry Lynd, and Stirling Bryan. Exploiting Order Effects to Improve the Quality of Decisions. *Patient Education and Counseling*, 96(2):197–203, aug 2014. ISSN 07383991. doi: 10.1016/j.pec.2014.05.021. URL `https://linkinghub.elsevier.com/retrieve/pii/S0738399114002262`.

[8] Joan C. Bartlett, Ilja Frissen, and Jamshid Beheshti. Comparing Academic and Everyday-Life Information Seeking Behavior Among Millennial Students. In *CHIIR 2020 - Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 402–406, New York, NY, USA, mar 2020. Association for Computing Machinery, Inc. ISBN 9781450368926. doi: 10.1145/3343413.3378006. URL `https://dl.acm.org/doi/10.1145/3343413.3378006`.

[9] W. Lance Bennett and Shanto Iyengar. A New Era of Minimal Effects? The Changing Foundations of Political Communication. *Journal of Communication*, 58(4):707–731, dec 2008. ISSN 00219916. doi: 10.1111/j.1460-2466.2008.00410.x. URL `https://academic.oup.com/joc/article/58/4/707/4098406`.

[10] John Bohannon. Mechanical Turk Upends Social Sciences. *Science*, 352(6291):1263–1264, jun 2016. ISSN 0036-8075. doi: 10.1126/science.352.6291.1263. URL `https://www.sciencemag.org/lookup/doi/10.1126/science.352.6291.1263`.

[11] Ye Cheny, Ke Zhouz, Yiqun Liuy, Min Zhangy, and Shaoping May. Meta-Evaluation of Online and Offlineweb Search Evaluation Metrics. In *SIGIR 2017 - Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 15–24, New York, NY, USA, aug 2017. Association for Computing Machinery, Inc. ISBN 9781450350228. doi: 10.1145/3077136.3080804. URL https://dl.acm.org/doi/10.1145/3077136.3080804.

[12] Joel Cooper and Kevin M. Carlsmith. Cognitive Dissonance. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, pages 76–78. Elsevier Inc., mar 2015. ISBN 9780080970875. doi: 10.1016/B978-0-08-097086-8.24045-2. URL https://collaborate.princeton.edu/en/publications/cognitive-dissonance.

[13] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. This Is Not What We Ordered: Exploring Why Biased Search Result Rankings Affect User Attitudes on Debated Topics. may 2021. doi: 10.1145/3404835.3462851. URL https://doi.org/10.1145/3404835.3462851.

[14] Alexey Drutsa, Gleb Gusev, and Pavel Serdyukov. Engagement Periodicity in Search Engine Usage: Analysis and Its Application to Search Quality Evaluation. In *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, pages 27–36, New York, New York, USA, feb 2015. Association for Computing Machinery, Inc. ISBN 9781450333177. doi: 10.1145/2684822.2685318. URL http://dl.acm.org/citation.cfm?doid=2684822.2685318.

[15] Georges Dupret and Mounia Lalmas. Absence Time and User Engagement. In *Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13*, pages 173–182, New York, New York, USA, 2013. ACM Press. ISBN 9781450318693. doi: 10.1145/2433396.2433418. URL https://doi.org/10.1145/2433396.2433418.

[16] Robert Epstein and Ronald E. Robertson. The Search Engine Manipulation Effect (Seme) and Its Possible Impact on the Outcomes of Elections. *Proceedings of the National Academy of Sciences*, 112(33):E4512–E4521, aug 2015. ISSN 0027-8424. doi: 10.1073/pnas.1419828112. URL https://doi.org/10.1073/pnas.1419828112.

[17] Robert Epstein, Ronald E. Robertson, David Lazer, and Christo Wilson. Suppressing the Search Engine Manipulation Effect (SEME). *Proceedings of the ACM on Human-Computer Interaction*, 1 (CSCW):1–22, nov 2017. ISSN 25730142. doi: 10.1145/3134677. URL https://dl.acm.org/doi/10.1145/3134677.

[18] James D Evans. *Straightforward Statistics for the Behavioral Sciences.* Thomson Brooks/Cole Publishing Co, Belmont, CA, US, 1996. ISBN 0-534-23100-4 (Hardcover).

[19] Franz Faul, Edgar Erdfelder, Albert Georg Lang, and Axel Buchner. G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. In *Behavior Research Methods*, volume 39, pages 175–191. Psychonomic Society Inc., 2007. doi: 10.3758/BF03193146. URL https://link.springer.com/article/10.3758/BF03193146.

[20] Leon Festinger. *A Theory of Cognitive Dissonance*. 2 edition, 1957. URL https://books.google.nl/books?hl=en&lr=&id=voeQ-8CASacC&oi=fnd&pg=PA1&dq=cognitive+dissonance&ots=9z55Prrhzy&sig=tmBNA_jEQ0erfXqJa7FvdgrkcnY&redir_esc=y#v=onepage&q=cognitivedissonance&f=false.

[21] P Fischer, A Kastenmüller, T Greitemeyer Journal of ..., and Undefined 2011. Threat and Selective Exposure: The Moderating Role of Threat and Decision Context on Confirmatory Information Search After Decisions. *Journal of Experimental Psychology: General*, 140(1):51–62, 2011. URL https://psycnet.apa.org/doi/10.1037/a0021595.

[22] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and Peter Holtz. Analyzing Knowledge Gain of Users in Informational Search Sessions on the Web. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval - CHIIR '18*, volume 2018-March, pages 2–11, New York, New

York, USA, feb 2018. ACM Press. ISBN 9781450349253. doi: 10.1145/3176349.3176381. URL http://dl.acm.org/citation.cfm?doid=3176349.3176381.

[23] Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. A Think-Aloud Study to Understand Factors Affecting Online Health Search. In *CHIIR 2020 - Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 273–282, New York, NY, USA, mar 2020. Association for Computing Machinery, Inc. ISBN 9781450368926. doi: 10.1145/3343413.3377961. URL https://dl.acm.org/doi/10.1145/3343413.3377961.

[24] Jeremy Goecks and Jude Shavlik. Learning Users' Interests by Unobtrusively Observing Their Normal Behavior. In *Proceedings of the 5th international conference on Intelligent user interfaces - IUI '00*, pages 129–132, New York, New York, USA, 2000. ACM Press. ISBN 1581131348. doi: 10.1145/325737.325806. URL http://portal.acm.org/citation.cfm?doid=325737.325806.

[25] Qi Guo, Ryen W White, Susan T Dumais, Jue Wang, and Blake Anderson. Predicting Query Performance Using Query, Result, and User Interaction Features. In *Adaptivity, personalization and fusion of heterogeneous information*, pages 198–201. 2010.

[26] Qi Guo, Dmitry Lagun, and Eugene Agichtein. Predicting Web Search Success With Fine-Grained Interaction Data. In *ACM International Conference Proceeding Series*, pages 2050–2054, New York, New York, USA, 2012. ACM Press. ISBN 9781450311564. doi: 10.1145/2396761.2398570. URL http://dl.acm.org/citation.cfm?doid=2396761.2398570.

[27] William Hart, Dolores Albarracín, Alice H Eagly, Inge Brechan, Matthew J Lindberg, and Lisa Merrill. Feeling Validated Versus Being Correct: A Meta-Analysis of Selective Exposure to Information. *Psychological bulletin*, 135(4):555, 2009. URL https://psycnet.apa.org/buy/2009-09537-004.

[28] Samuel Ieong, Nina Mishra, Eldar Sadikov, and Li Zhang. Domain Bias in Web Search. In *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*, page 413, New York, New York, USA, 2012. ACM Press. ISBN 9781450307475. doi: 10.1145/2124295.2124345. URL http://dl.acm.org/citation.cfm?doid=2124295.2124345.

[29] Jiepu Jiang, Daqing He, and James Allan. Searching, Browsing, and Clicking in a Search Session: Changes in User Behavior by Task and Over Time. In *SIGIR 2014 - Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 607–616, New York, New York, USA, 2014. Association for Computing Machinery. ISBN 9781450322591. doi: 10.1145/2600428.2609633. URL http://dl.acm.org/citation.cfm?doid=2600428.2609633.

[30] Daniel Kahneman. A Perspective on Judgment and Choice: Mapping Bounded Rationality. *American Psychologist*, 58(9):697–720, 2003. ISSN 1935-990X. doi: 10.1037/0003-066X.58.9.697. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.58.9.697.

[31] Rishita Kalyani and Ujwal Gadiraju. Understanding User Search Behavior Across Varying Cognitive Levels. In *HT 2019 - Proceedings of the 30th ACM Conference on Hypertext and Social Media*, pages 123–132. Association for Computing Machinery, Inc, sep 2019. ISBN 9781450368858. doi: 10.1145/3342220.3343643.

[32] Markus Kattenbeck and David Elsweiler. Understanding Credibility Judgements for Web Search Snippets. *Aslib Journal of Information Management*, 71(3):368–391, may 2019. ISSN 2050-3806. doi: 10.1108/AJIM-07-2018-0181. URL https://www.emerald.com/insight/content/doi/10.1108/AJIM-07-2018-0181/full/html.

[33] Mark T. Keane, Maeve O'Brien, and Barry Smyth. Are People Biased in Their Use of Search Engines? *Communications of the ACM*, 51(2):49–52, feb 2008. ISSN 0001-0782. doi: 10.1145/1314215.1314224. URL https://dl.acm.org/doi/10.1145/1314215.1314224.

[34] Diane Kelly and Leif Azzopardi. How Many Results per Page? A Study of SERP Size, Search Behavior and User Experience. In *SIGIR 2015 - Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 183–192, New York, New York, USA, aug 2015. Association for Computing Machinery, Inc. ISBN 9781450336215. doi: 10.1145/2766462.2767732. URL http://dl.acm.org/citation.cfm?doid=2766462.2767732.

[35] Joshua Klayman and Young-won Ha. Confirmation, Disconfirmation, and Information in Hypothesis Testing. *Psychological Review*, 94(2):211–228, 1987. ISSN 0033-295X. doi: 10.1037/0033-295X.94.2.211. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.94.2.211.

[36] Bart P. Knijnenburg, Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5):441–504, oct 2012. ISSN 09241868. doi: 10.1007/s11257-011-9118-4. URL https://link.springer.com/article/10.1007/s11257-011-9118-4.

[37] Silvia Knobloch-Westerwick, Benjamin K. Johnson, and Axel Westerwick. Confirmation Bias in Online Searches: Impacts of Selective Exposure Before an Election on Political Attitude Strength and Shifts. *Journal of Computer-Mediated Communication*, 20(2):171–187, mar 2015. ISSN 10836101. doi: 10.1111/jcc4.12105. URL https://academic.oup.com/jcmc/article/20/2/171/4067554.

[38] Annie Y.S. Lau and Enrico W. Coiera. Do People Experience Cognitive Biases while Searching for Information? *Journal of the American Medical Informatics Association*, 14(5):599–608, sep 2007. ISSN 10675027. doi: 10.1197/jamia.M2411. URL https://academic.oup.com/jamia/article/14/5/599/721400.

[39] Dirk Lewandowski. Users' Understanding of Search Engine Advertisements. *Journal of Information Science Theory and Practice*, 5(4):6–25, 2017. ISSN 22874577. doi: 10.1633/JISTaP.2017.5.4.1. URL https://doi.org/10.1633/JISTaP.2017.5.4.1.

[40] Dirk Lewandowski, Friederike Kerkmann, Sandra Rümmele, and Sebastian Sünkler. An Empirical Investigation on Search Engine Ad Disclosure. *Journal of the Association for Information Science and Technology*, 69(3):420–437, mar 2018. ISSN 23301635. doi: 10.1002/asi.23963. URL http://doi.wiley.com/10.1002/asi.23963.

[41] Mengyang Liu, Jiaxin Mao, Yiqun Liu, Min Zhang, and Shaoping Ma. Investigating Cognitive Effects in Session-level Search User Satisfaction. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, volume 19, pages 923–931, New York, NY, USA, jul 2019. Association for Computing Machinery. ISBN 9781450362016. doi: 10.1145/3292500.3330981. URL https://dl.acm.org/doi/10.1145/3292500.3330981.

[42] Yiqun Liu, Ye Chen, Jinhui Tang, Jiashen Sun, Min Zhang, Shaoping Ma, and Xuan Zhu. Different Users, Different Opinions: Predicting Search Satisfaction With Mouse Movement Information. In *SIGIR 2015 - Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 493–502, New York, NY, USA, aug 2015. Association for Computing Machinery, Inc. ISBN 9781450336215. doi: 10.1145/2766462.2767721. URL https://dl.acm.org/doi/10.1145/2766462.2767721.

[43] Carla Teixeira Lopes and Edgar Ramos. Studying How Health Literacy Influences Attention During Online Information Seeking. In *CHIIR 2020 - Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 283–291, New York, NY, USA, mar 2020. Association for Computing Machinery, Inc. ISBN 9781450368926. doi: 10.1145/3343413.3377966. URL https://dl.acm.org/doi/10.1145/3343413.3377966.

[44] Lori Lorigo, Maya Haridasan, Hrönn Brynjarsdóttir, Ling Xia, Thorsten Joachims, Geri Gay, Laura Granka, Fabio Pellacini, and Bing Pan. Eye Tracking and Online Search: Lessons Learned and Challenges Ahead. *Journal of the American Society for Information Science and Technology*, 59 (7):1041–1052, may 2008. ISSN 15322882. doi: 10.1002/asi.20794. URL http://doi.wiley.com/10.1002/asi.20794.

[45] David Maxwell and Claudia Hauff. LogUI : Contemporary Logging Infrastructure for Web-Based Experiments. In *ECIR*, pages 525–530, 2021.

[46] Panagiotis Takis Metaxas and Yada Pruksachatkun. Manipulation of Search Engine Results during the 2016 US Congressional Elections, 2017. URL https://repository.wellesley.edu/islandora/object/ir%3A264/.

[47] Ben Miroglio, David Zeber, Jofish Kaye, and Rebecca Weiss. The Effect of Ad Blocking on User Engagement with the Web. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, pages 813–821, New York, New York, USA, apr 2018. ACM Press. ISBN 9781450356398. doi: 10.1145/3178876.3186162. URL http://dl.acm.org/citation.cfm?doid=3178876.3186162.

[48] Vidhya Navalpakkam, Ladawn Jentzsch, Rory Sayres, Sujith Ravi, Amr Ahmed, and Alex Smola. Measurement and Modeling of Eye-Mouse Behavior in the Presence of Nonlinear Page Layouts. In *WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web*, pages 953–963. Association for Computing Machinery, 2013. ISBN 9781450320351. doi: 10.1145/2488388.2488471. URL http://googleblog.blogspot.com/2012/05/introducing-.

[49] Geoff Norman. Likert Scales, Levels of Measurement and the "Laws" of Statistics. *Advances in Health Sciences Education*, 15(5):625–632, dec 2010. ISSN 1382-4996. doi: 10.1007/s10459-010-9222-y. URL http://link.springer.com/10.1007/s10459-010-9222-y.

[50] Alamir Novin and Eric Meyers. Making Sense of Conflicting Science Information: Exploring Bias in the Search Engine Result Page. In *CHIIR 2017 - Proceedings of the 2017 Conference Human Information Interaction and Retrieval*, pages 175–184, New York, NY, USA, mar 2017. Association for Computing Machinery, Inc. ISBN 9781450346771. doi: 10.1145/3020165.3020185. URL https://dl.acm.org/doi/10.1145/3020165.3020185.

[51] Heather L. O'Brien, Paul Cairns, and Mark Hall. A Practical Approach to Measuring User Engagement With the Refined User Engagement Scale (Ues) and New Ues Short Form. *International Journal of Human-Computer Studies*, 112:28–39, apr 2018. ISSN 10715819. doi: 10.1016/j.ijhcs.2018.01.004. URL https://linkinghub.elsevier.com/retrieve/pii/S1071581918300041.

[52] Heather L. O'Brien, Jaime Arguello, and Rob Capra. An Empirical Study of Interest, Task Complexity, and Search Behaviour on User Engagement. *Information Processing & Management*, 57(3):102226, may 2020. ISSN 03064573. doi: 10.1016/j.ipm.2020.102226. URL https://linkinghub.elsevier.com/retrieve/pii/S0306457319301591.

[53] Stefan Palan and Christian Schitter. Prolific.ac—A Subject Pool for Online Experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, mar 2018. ISSN 22146350. doi: 10.1016/j.jbef.2017.12.004. URL https://linkinghub.elsevier.com/retrieve/pii/S2214635017300989.

[54] Srishti Palani, Adam Fourney, Shane Williams, Kevin Larson, Irina Spiridonova, and Meredith Ringel Morris. An Eye Tracking Study of Web Search by People with and Without Dyslexia. In *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 729–738, New York, NY, USA, jul 2020. Association for Computing Machinery, Inc. ISBN 9781450380164. doi: 10.1145/3397271.3401103. URL https://dl.acm.org/doi/10.1145/3397271.3401103.

[55] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated Communication*, 12(3):801–823, apr 2007. ISSN 10836101. doi: 10.1111/j.1083-6101.2007.00351.x. URL https://academic.oup.com/jcmc/article/12/3/801-823/4582975.

[56] Frances A. Pogacar, Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. The Positive and Negative Influence of Search Results on People's Decisions About the Efficacy of Medical Treatments. In *ICTIR 2017 - Proceedings of the 2017 ACM SIGIR International Conference on the Theory of Information Retrieval*, pages 209–216, New York, NY, USA, oct 2017. Association for Computing Machinery, Inc. ISBN 9781450344906. doi: 10.1145/3121050.3121074. URL https://dl.acm.org/doi/10.1145/3121050.3121074.

[57] Suppanut Pothirattanachaikul, Yusuke Yamamoto, Takehiro Yamamoto, and Masatoshi Yoshikawa. Analyzing the Effects of Document's Opinion and Credibility on Search Behaviors and Belief Dynamics. In *International Conference on Information and Knowledge Management, Proceedings*, pages 1653–1662, New York, NY, USA, nov 2019. Association for Computing Machinery. ISBN 9781450369763. doi: 10.1145/3357384.3357886. URL https://dl.acm.org/doi/10.1145/3357384.3357886.

[58] Suppanut Pothirattanachaikul, Takehiro Yamamoto, Yusuke Yamamoto, and Masatoshi Yoshikawa. Analyzing the Effects of ”People Also Ask” on Search Behaviors and Beliefs. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media, HT 2020*, pages 101–110, New York, NY, USA, jul 2020. Association for Computing Machinery, Inc. ISBN 9781450370981. doi: 10.1145/3372923.3404786. URL https://dl.acm.org/doi/10.1145/3372923.3404786.

[59] Emily Pronin, Daniel Y. Lin, and Lee Ross. The Bias Blind Spot: Perceptions of Bias in Self Versus Others, jul 2002. ISSN 01461672. URL https://journals.sagepub.com/doi/abs/10.1177/0146167202286008.

[60] Ronald E. Robertson, David Lazer, and Christo Wilson. Auditing the Personalization and Composition of Politically-Related Search Engine Results Pages. In *The Web Conference 2018 - Proceedings of the World Wide Web Conference, WWW 2018*, pages 955–965, New York, New York, USA, apr 2018. Association for Computing Machinery, Inc. ISBN 9781450356398. doi: 10.1145/3178876.3186143. URL http://dl.acm.org/citation.cfm?doid=3178876.3186143.

[61] Kerry Rodden and Xin Fu. Exploring How Mouse Movements Relate to Eye Movements on Web Search Results Pages. Technical report, 2007. URL http://www.rodden.org/kerry/publications/wisi07/.

[62] Sebastian Schultheiß and Dirk Lewandowski. How Users' Knowledge of Advertisements Influences Their Viewing and Selection Behavior in Search Engines. *Journal of the Association for Information Science and Technology*, 72(3):285–301, mar 2021. ISSN 2330-1635. doi: 10.1002/asi.24410. URL https://onlinelibrary.wiley.com/doi/10.1002/asi.24410.

[63] Amos Tversky and Daniel Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, sep 1974. ISSN 0036-8075. doi: 10.1126/science.185.4157.1124. URL https://www.sciencemag.org/lookup/doi/10.1126/science.185.4157.1124.

[64] Andre Calero Valdez, Martina Ziefle, and Michael Sedlmair. Priming and Anchoring Effects in Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):584–594, jan 2018. ISSN 1077-2626. doi: 10.1109/TVCG.2017.2744138. URL http://ieeexplore.ieee.org/document/8022891/.

[65] Ryen W. White. Beliefs and Biases in Web Search. In *SIGIR 2013 - Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3–12, New York, New York, USA, 2013. ACM Press. ISBN 9781450320344. doi: 10.1145/2484028.2484053. URL http://dl.acm.org/citation.cfm?doid=2484028.2484053.

[66] Ryen W White and Ahmed Hassan. Content Bias in Online Health Search. *ACM Transactions on the Web (TWEB)*, 8(4):1–33, nov 2014. doi: 10.1145/2663355. URL http://dx.doi.org/10.1145/2663355.

[67] Luyan Xu, Xuan Zhou, and Ujwal Gadiraju. How Does Team Composition Affect Knowledge Gain of Users in Collaborative Web Search? In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, pages 91–100, New York, NY, USA, jul 2020. ACM. ISBN 9781450370981. doi: 10.1145/3372923.3404784. URL https://dl.acm.org/doi/10.1145/3372923.3404784.

[68] Luyan Xu, Mengdie Zhuang, and Ujwal Gadiraju. How Do User Opinions Influence Their Interaction With Web Search Results? In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pages 240–244, New York, NY, USA, jun 2021. ACM. ISBN 9781450383660. doi: 10.1145/3450613.3456824. URL https://dl.acm.org/doi/10.1145/3450613.3456824.

[69] Takehiro Yamamoto, Yusuke Yamamoto, and Sumio Fujita. Exploring People's Attitudes and Behaviors Toward Careful Information Seeking in Web Search. In *International Conference on Information and Knowledge Management, Proceedings*, pages 963–972, New York, NY, USA, oct 2018. Association for Computing Machinery. ISBN 9781450360142. doi: 10.1145/3269206.3271799. URL https://dl.acm.org/doi/10.1145/3269206.3271799.

[70] Yusuke Yamamoto and Takehiro Yamamoto. Query Priming for Promoting Critical Thinking in Web Search. In *CHIIR 2018 - Proceedings of the 2018 Conference on Human Information Interaction and Retrieval*, volume 2018-March, pages 12–21, New York, New York, USA, feb 2018. Association for Computing Machinery, Inc. ISBN 9781450349253. doi: 10.1145/3176349.3176377. URL http://dl.acm.org/citation.cfm?doid=3176349.3176377.

[71] Yusuke Yamamoto, Hiroaki Ohshima, Takehiro Yamamoto, and Hiroshi Kawakami. Web Access Literacy Scale to Evaluate How Critically Users Can Browse and Search for Web Information. In *WebSci 2018 - Proceedings of the 10th ACM Conference on Web Science*, volume 10, pages 97–106, New York, NY, USA, may 2018. Association for Computing Machinery, Inc. ISBN 9781450355636. doi: 10.1145/3201064.3201072. URL https://dl.acm.org/doi/10.1145/3201064.3201072.

[72] Yan Zhang and Shijie Song. Older Adults' Evaluation of the Credibility of Online Health Information. In *CHIIR 2020 - Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 358–362, New York, NY, USA, mar 2020. Association for Computing Machinery, Inc. ISBN 9781450368926. doi: 10.1145/3343413.3377997. URL https://dl.acm.org/doi/10.1145/3343413.3377997.

[73] Yan Zhang, Yalin Sun, and Bo Xie. Quality of Health Information for Consumers on the Web: A Systematic Review of Indicators, Criteria, Tools, and Evaluation Results. *Journal of the Association for Information Science and Technology*, 66(10):2071–2084, oct 2015. ISSN 23301635. doi: 10.1002/asi.23311. URL http://doi.wiley.com/10.1002/asi.23311.