How Is Attitude Change in Web Search Influenced by Users' Informedness?

S.A.N. Kulane

MSc Thesis

TU Delft

How Is Attitude Change in Web Search Influenced by Users' Informedness?

MSc Thesis

by

S.A.N. Kulane

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Friday April 14, 2023 at 11:00 AM.

Student number: 4319540

Thesis committee:	Prof. dr. ir. G. J. P. M. Houben,	TU Delft, chair
	Dr. U. K. Gadiraju,	TU Delft, supervisor
	A. Rieger,	TU Delft, co-supervisor
	Dr. P. K. Murukannaiah,	TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Preface

This thesis is the final chapter of my education at TU Delft. It concludes my curriculum for the MSc Computer Science at EEMCS faculty. It has been a period that I look back on with many memories.

I want to thank both Alisa and Uwjal. Thanks to their invaluable input, discussions and guidance, I learned a lot regarding the topic of my thesis and more. I would not be aware of this topic if it were not for their introduction. Thank you for your support and encouragement.

The majority of my master's education was during the pandemic. Therefore I would like to thank my family and friends for their support during this period. Without them, it would have been much more challenging to finish my master's.

S.A.N. Kulane
Delft, March 2013

Abstract

Search engines are used to gather and collect information. This interaction sometimes influences the user and changes their attitude towards a topic after such interaction. Prior work has shown that it is a complex endeavour to understand attitude change, as there are many things that can influence a user during a search session. One of the aspects that have not been researched is the influence of informedness. Therefore we examine the role of informedness in attitude change. To do that, we first defined what a well-informed user in the context of web search is. Then a user study is conducted to understand informedness's role in attitude change. From the platform Prolific, we recruited participants (N=320) for the main user study. The experiment is a 3 x 3 (Informedness level x SERP viewpoint bias) factorial between-subjects study and it requires interaction with the provided search results. The participants were placed in one of three categories, ”Well Informed”, ”Mildly Informed”, and ”Uninformed”, based on their results on a knowledge questionnaire. We created 3 different conditions for the search engine results page (SERP). The SERP had a viewpoint-biased ranking that was either supportive, opposing or balanced towards the topic. There were two topics, abortion and obesity, to which the participants were assigned. Our findings showed that 37% of the participants with a valid submission had changed attitudes. The findings from the user study suggest that the level of informedness does not play a role in facilitating attitude change among users during a web search. This has implications for the design of information retrieval systems and web search experiences in general.

Contents

List of Figures

List of Tables

1 Introduction

Search engines are used frequently to find and collect information. How people use and are influenced by search engine results pages has been a topic of much research. Yet understanding why users change their attitude is important, for example, to detect and avoid influences in election processes [9], or to better understand why some are changing or not changing their attitude towards vaccines. Existing research has shown that users can be affected to the point of changing their attitude on vaccinations by the quality of the results [1]. However, does this imply that higher quality or more knowledgeable results are enough to lead to an attitude change? We do not know yet. So far, the dynamics of attitude change during web searches have not been fully understood. Prior research has found some aspects that might influence it, for example, the role that a weak prior attitude and strong attitude play in attitude change [36]. White [36] found that users with weaker attitudes are more likely to change their attitudes. In contrast, those with strong attitudes are unlikely to change their attitude [36]. Why this is the case is not clear, this could be related to finding more information, gaining more knowledge and thus re-evaluating prior attitudes.

We come across the notion of a well-informed user in the literature regarding attitudes and knowledge. This ties closely with another definition of justified belief, where the opinion or attitude held by an individual must meet certain requirements to be considered justified or responsible. One such requirement is a high level of knowledge to form a justified belief or attitude [27, 12]. Existing work on attitude change in web search does not look at this aspect. It stays focused on the change in attitude and the effect web search results might have. Yet our aim is to look at the prior foundation of the attitude before a search session and investigate differences in attitude change during web searches between more and less informed users.

The aim of this thesis is to add more knowledge to the growing literature on search as opinion formation by addressing this gap. We focus on prior informedness and its role in attitude change during web search. A better understanding of the role of informedness on attitude change advances our knowledge towards building search engines that can support users in forming justified or well-informed attitudes. This thesis aims to address this gap and contribute to understanding this facet of web search interaction.

1.1. Research questions

For this thesis, we formulated the following overarching question: **How is attitude change in web search influenced by users' informedness?** To address the gap in research on attitude change and to understand the role of informedness, we have formulated two research questions. The first one is to define a well-informed user. Then we look into the effect of different levels of informed users on attitude change.

RQ1 What is a well-informed user in the context of web search?

RQ2 Are well-informed users less susceptible to attitude change than uninformed or mildly informed

users in biased search?

To define well-informed users, we will conduct a literature review of different fields and synthesise a definition applicable to the web search context. This definition will then be applied to measure informedness in a 3x3 factorial experiment that will be conducted to investigate and answer RQ2. In the experiment, we will control the viewpoint bias on the search engine results page (SERP) and the informedness of the participants. The participants will be divided into three categories of informedness levels. They are well-informed, mildly informed and uninformed. The SERP viewpoint bias will be implemented using a viewpoint-biased ranking of the results, where the viewpoint bias of the SERP will be ranked higher. There are three options for the SERP. It is either biased in favour, biased against or balanced. By controlling both variables, we can measure the effect on attitude change with minimal interference from other SERP interactions. Our results suggest that informedness's effect does not seem to influence attitude change. However, our definition of well-informed users did hold up and shows promise if behavioural characteristics are added. The definition has two main requirements; one is knowledge of a topic and the other is being aware of other stances and attitudes.

1.2. Contribution

There are several contributions that this thesis will make

- A definition of well-informed users in the context of web search
- A preregistered user study of a 3x3 factorial experiment design. It will be used to understand the role of informedness in web search [1].
- A data set with the participants' questionnaire responses and their SERP interaction. The data obtained from all experiments will be made public and available. This includes two smaller experiments to gather data and the main user study [2].
- The source code for the custom SERP page used in the experiment will be made public. A readme file with additional implementation details will be included as well.[2]

1.3. Outline

In the next chapter, Chapter 2, we will discuss and present a summary of related work on five topics. We will look into information gathering, search as learning, search biases, attitude change and emotions in search. They are needed to understand better the topic of attitude change and several related ongoing research fields. Chapter 3 will look into literature outside of web search to understand how informedness is defined and tested in other domains. This will be crucial to developing a working definition used in user experiments. In Chapter 4, the experiment and required materials are discussed. Here the choices made in experiment design, technical details and more are explained and made clear. All three experiments will be discussed one by one. The first two are to gather the required data for the main user study, such as questions and viewpoint annotations. The results of hypotheses testing and exploratory analyses will be presented in Chapter 5. The implications and limitations of the results will be discussed in Chapter 6. The suggestions for future work and conclusion will be given in Chapter 7.

[1] https://doi.org/10.17605/OSF.IO/G3R5Z
[2] https://doi.org/10.17605/OSF.IO/YFX6J

2

Related Work

In this chapter, we look into the research that has already been done related to our topic. In Section 2.1, we look at the research regarding information seeking and web search. The differences in behaviours and methods of how information is gathered will be discussed using existing research. The next Section 2.2 will look at learning from search. This is to understand how learning new information and knowledge gain happens in search. It plays a role in our ability to understand the informedness of search users. After this, we will discuss search biases in Section 2.3. This will increase our understanding of how biases have an effect on the search behaviour of the user. Then we review the research regarding attitude change in web search in Section 2.4. Since the topic of this thesis is attitude change we need to analyse the existing research regarding attitude change to understand what is known and unknown. We will end the literature overview with a look at emotion and its role in web search in Section 2.5.
In this chapter, we will use several terminologies which refer to the same things. One such example is belief change and attitude change. Both terms are used in the literature. To give an accurate representation both will be used. However, after this chapter, preference will be given to attitude change.

2.1. Information seeking

2.1.1. Search behaviour

The way users conduct searches and how different aspects influence their search interaction has been researched before. The differences in search behaviour have been found and observed in several works. The research done by Hölscher and Strube [17] shows that there is a behavioural difference between novice and experienced users of search. Their results showed that the group with no domain or search experience, the double novice, used the most query reformulations compared to the other groups. The changes were small and insignificant, leading to multiple reformulations. Furthermore, they also visited the fewest result in comparison as well. Those that they did visit were often irrelevant. In contrast, the expert group with domain and search knowledge barely used the 'back' button to return to previous results, something that the authors noticed to be common with the less experienced groups. The groups with some knowledge in either search or topic used it to compensate for the other domain they lacked. Those with lower levels of knowledge had less flexibility in their strategies during the search [17]. The differences in query formulations and vocabulary between experts and novices have also been observed through a log study by White, Dumais, and Teevan [37].

2.1.2. Critical information seeking

The aim of using search engines is often to find information. Research by Yamamoto, Yamamoto, and Fujita [40] aims to understand users' attitudes towards critical, careful information gathering. This is defined by the ability to define the information need, source, evaluate and use the information [40]. Their study considers four types of thinking styles: need for cognition, flexible thinking, faith in intuition and general trust. The need for cognition relates to enjoying effortful cognitive tasks. Flexible thinking is the ability to change the belief in the face of evidence, while faith in intuition relies on intuition for making a decision. The fourth, general trust, relates to a person's general honesty and trust [40]. The hypothesis of the study is that some thinking styles have a positive correlation to verification attitudes, while other

thinking styles have a negative correlation. Verification attitudes are attitudes of users willing to put in an additional effort to verify the sources, such as looking if the source is up to date [40]. The authors gathered and analysed the query logs of a search engine for a month, with the consent of the users. They found that users with critical attitudes exhibit different search behaviour. Some of the behaviour included complex search methods and longer queries, but those sessions were shorter as they found the information. The behaviour could also be attributed to domain knowledge [40, 37]. The result also showed that those who ranked high on verification attitude clicked lower ranking results and used verification terms such as research, proof, survey and comparison in their query. Another important finding is that the differences in behaviours were not based on educational background. However, a high education background did lead to a likelihood of high verification attitude. The researchers explain this difference as an insufficient attitude, even though the educational background does provide them with the skills to do so. This follows from their conclusion of the study as well. The verification attitude of the users is positively correlated to the need for thinking, education and search expertise [40].

Yamamoto and Yamamoto [42] conducted another user study to investigate increasing critical thinking in web searches. The method used to promote critical thinking was query priming. Query priming was done by suggesting query completions that stimulate, and prime, the user for critical thinking [42]. Words like research, validation, data and comparison can be used for query priming. They conducted a user study in which they analysed the search behaviour and questionnaire responses. From the results, they found that the priming UI (see Figure 2.1) resulted in more queries and SERP (search engine results page) visits. Furthermore, education levels only affected the number of queries since the university-educated participants issued more queries due to priming UI. The result of the study leads to the conclusion that the priming UI works. University-educated participants collected more evidence with valid references [42]. However, there is little to support that this change in behaviour will be sustained without the priming effect [42].

Figure 2.1: Priming UI on the left. Image is from Yamamoto and Yamamoto [42]

In earlier work by Yamamoto and Shimada [41], it was found that disputed topic suggestions lead to more time spent on the SERP and, more importantly, more consideration of the credibility of the results [41]. Disputed topics in their work refer to results claimed to be suspicious irrespective of the truth.

There is also research done into the effect of bad results. In the context of medical search, this can lead to dangerous situations. Pogacar et al. [28] showed that search results with a bias towards correct results were shown users had an accuracy of 65%. With negative and wrong information results, it dropped from 43% to 23%. This has a real-life impact on the decisions of users. The authors warn that bad results not only take time but also damage the decision of users. In conclusion, the authors remarked that search engines need to actively support users in finding the correct information.

2.2. Search as learning

Search engines are not just tools to gather information, but also tools to learn and get educated (search as learning). The two, information seeking and learning can be linked by using the framework of human cognition [14]. In fact, Ghosh, Rath, and Shah [14] make an even bolder statement: "...learning is an outcome of information seeking." [14]. Their research investigates the relationship between learning and searching by designing a user study. The tasks were along the levels of cognitive complexity. The levels are: remember, understand, apply, analyze, evaluate and create [4]. The first two levels,

remembering and understanding, were combined in the first task. The level create was not included in the experiment. The results show that the highest amount of search results visited was for the analysis task, and the second spot was for remembering and understanding the task. The results also showed that the users gained knowledge after the search task, nearly 35% increase in topic knowledge over all four tasks. However, there was also a significant difference in the knowledge of the four cognitive complexity levels. In each of the four cognitive levels, they observed differences in search behaviours as well.

Another paper in this field by Gadiraju et al. [13] also analyzed the knowledge gained in web searches. They created a knowledge questionnaire for ten topics and measured the gain in knowledge. From the results, they found that the highest gain in knowledge occurred in topics with the least familiarity for the participants. The participants were found to have used longer search queries at the end of the session. This behaviour can be linked to knowledge gain since the queries are more advanced. Overall on average, the researchers found a knowledge increase of nearly 20% and 70% of the participants showed signs of knowledge increase.

The knowledge gained in web searches in other settings has been researched as well. For example, the research by Xu, Zhou, and Gadiraju [38] examined the knowledge gained in collaborative web searches. The term *collaborative search* describes a search where users or participants work together to gather information. The researchers looked at the role of the collaboration but also the effect of the user's characteristics. The results of their model gave knowledge and education alongside other descriptions as predictors for knowledge gain. The analysis of the data showed that there was a 44% increase in knowledge gain. This is higher than previous research by Gadiraju et al. [13] where the knowledge gain for single user search was 35%. Looking at the role of domain knowledge and knowledge gain, they found out that a low domain knowledge but high education (e.g. college) resulted in a higher knowledge gain than users with lower education levels. However, if users had a high or average (moderate) domain knowledge, those with lower education had a higher knowledge gain than higher-educated users. They showed that knowledge and education levels have an effect on the knowledge gain of the users.
One of the works done by other researchers later looked into a more dynamic approach where the search engine considers the users' knowledge. This was shown in the research by El Zein and da Costa Pereira [8], where they proposed and tested a framework for information retrieval that takes users into account. It kept track of the users' progress throughout the interaction. Thereby providing results that fit the user. The result was an overall knowledge gain for the participants.
Other works have looked at the capturing and measuring of knowledge in search. While questionnaires are the status quo, Yu et al. [43] aimed to automate knowledge measurements by means of prediction. They concluded that predicting knowledge gain can be done. Even the knowledge state can be predicted using the search behaviour. The prediction performs better with either low or high knowledge state or gain.

2.3. Search bias

A large amount of research has been done into the biases that are present in the search. Those biases can be introduced through different methods. It could be due to the search engine or the user. Both types of biases impact the user interaction with the results in the search engine. The huge variation in biases makes it challenging to cover every single type of bias, therefore a selection and categorization of biases that impact users' behaviour have been chosen to highlight to understand the effect they can have on this research in search. The chosen selection of biases is something that we might encounter in our experimentation and is therefore important to understand.

2.3.1. Cognitive bias

Cognitive biases can have a negative or positive effect on decisions making such as the gathering of information in search [2]. Cognitive biases come in various types, Azzopardi [2] has categorized into four categories. The four major categories are *(1) Information overload, (2) Information scarcity, (3) Urgency and (4) Information retention* [2]. The following selection of biases has been identified in work done by Azzopardi [2].

1. Availability bias: Choosing an answer or stance based on availability and how easily it can be found.
2. Framing effect: Decisions are influenced by the method of presentation of information.
3. Anchoring bias: Making decisions based on first information obtained or first impressions.
4. Confirmation bias: Tendency to choose the information that matches prior knowledge or stance. This includes the rejection of conflicting information.
5. Reinforcement effects: Decision is influenced by being stimulated with the same information or stance frequently.
6. Ambiguity effects: Avoiding results which may cause uncertainty, even if it is favourable [2]. In search, this can be manifested by choosing known sources over unknown sources.
7. Priming effects: Users are exposed to stimuli that influence their choice. This could be done by images or text.
8. Order effects: The order in which information is provided can influence the users' choice.

Plenty of other biases can have an impact on a user's search experience. This is a short selection of biases. For example, *authority bias* is one where the user trusts the results due to the source. This could be the search engine that provides the ranking, position bias [2], or the domain page of the results. The latter is also called *domain bias* [18]. The biases identified by prior research do not have to occur in isolation, multiple biases can be present in a session. For example, anchoring and priming bias often occur together as both have the first results as a source of bias [2]. Even the stance of the results can influence the user. If the results are formulated positively or negatively, it can affect the user as well, a study showed that users preferred results with a positive emotion [19].
Novin and Meyers [26] looked into the effect of biases on the SERP. They looked into biases such as priming, anchoring, framing and availability heuristics. They found a priming effect where users rated known sources, such as Wikipedia, higher than unknown sources, such as academic papers. The anchoring effect played a role when high-ranked results were trusted more than others, even though the information was one-sided. The framing effect in SERP can be observed when conflicting results are ranked lower than others. Finally, they found in their study that if a result is out of place it will be deemed less useful by users [26]. The authors suggest being more transparent about why results are ranked and identifying their connections, e.g. other viewpoints. Another suggestion is to add results that cover the breadth of the topic before going into depth. This will provide better context to the users.

2.3.2. Confirmation bias

In the study done by Xu, Zhuang, and Gadiraju [39], they looked at users with strong opinions and actions during a search session. The goal of this research was the intent of the search, purposeful and purposeless. The resulting user study showed that users with purposeless were easier influenced by the search results. Users with strong supportive opinions would have more engagement with a topic, such as more clicks and spending more time. Thirdly, they found that users tend to believe results that align with their opinions [39]. The last result is also known as confirmation bias.
More research on the impact of confirmation bias on the behaviour of users in web search is done by Suzuki and Yamamoto [33]. Their research used a user study where users had prior beliefs. The users were introduced to health-related topics before executing a search task. They showed that prior belief impacted the behaviour since users did not look further in the results list than the top results. They chose the results that matched their prior belief. They concluded that confirmation bias mitigates users' health literacy.

2.4. Attitude change in search

Part of the research into user interaction and search engines relates to the changes in the attitudes of users. White [35] did a retrospective survey where users were asked to reflect on a past search experience. They were asked what they believed before and after the search. The focus was on the

changing beliefs due to search. The results showed that neutral users were more likely to move towards a positive attitude. Users that had a strong attitude were more inclined to confirmatory results. A later study by White [36] concluded that strong opinions are unlikely to change, but those with weaker beliefs may be more open to considering alternative results. Furthermore, search manipulation to alter belief is only effective on those with a weaker stance [36].

Allam, Schulz, and Nakamoto [1] did research into attitude change, specifically on the topic of vaccines. Their user study looked at the effect of ranking on users' attitudes. They had two hypotheses one regarding the knowledge gain users can get from quality results and the other one regarding the message of the result. For these hypotheses, they conducted two experiments with similar setups. The results indicated that their first hypothesis was correct. High-quality results lead to knowledge gain. Users with non-quality results did not show a gain in knowledge. The same group (high quality) was the only group to show positive attitude change. No other group had the same result. They concluded that users are not able to effectively recognize high-quality results, which suggests that users are bad at recognizing results that mislead them [1].

A user study by Draws et al. [7] looked at attitude changes in web searches. They aimed to understand the search engine manipulation effect (SEME). SEME is the type of attitude change caused by viewing biased search results [7]. The reason for this is not understood. In this user study, one of the two biases that are expected to play a role is examined, namely ordering effects. Their findings show no evidence for ordering effects. Furthermore, their result suggests that the different levels of biased results were not detected by the participants. Yet a majority of the participants experienced an attitude change, nearly 70% of the participants. Most of them (57%) reported an attitude that was more supportive of the topic. From the exploratory analysis, they found that exposure effects could play a role in this type of attitude change.

Pothirattanachaikul et al. [29] Looked at the effect of result credibility on users' behaviour and belief dynamics. From their experiment, they found that results with high credibility and consistency with the users' beliefs meant that users were more likely to keep their original beliefs. However, results that were inconsistent with the prior belief led to more search queries being used. The final conclusion was that users tend to change their beliefs if they came across results that were not aligned with their prior beliefs. They did not find any relation between belief change and search behaviour.

Roscoe et al. [30] looked into the effect of the stance of the webpage on the user. In their user study, they used the topic of water bottles. They found that users exposed to searches with a positive overall stance were more likely to buy a water bottle than users exposed to negative searches.

Several of the existing research shows that users are influenced during a search interaction. This leads to changes in attitude, which has been observed and documented multiple times. Yet the underlying reason has not been understood. We know elements in search, such as the stances and qualities of the results, do impact the user. However, this only impacts part of the users, not all of them. To understand why this impacts some and not all, we will look into the role of informedness.

2.5. Emotion

As part of this thesis, we are also looking into the emotional response caused by an attitude change. The reason is that cognitive dissonance can cause negative emotions [15, 11]. This occurs when the knowledge of the user and what is in front of them don't match. The negative emotions can be a motivating source to change belief and accept the dissonant information [15]. As was mentioned earlier in the section above by Pothirattanachaikul et al. [29] users changed beliefs when they came across information conflicting with their initial stance. The role and effect of emotion is an active research field. Some of the research in that field relates to understanding our research in attitude change.

Kazai, Thomas, and Craswell [19] looked at the emotions of results in the SERP. They wanted to understand the role of emotion in search and as a result, the decisions informed by search. They gathered data on several topics, both positive and negative emotions. They concluded, through analyzing the search and click data, that clicked results were more often positive compared to results that are not clicked [19]. Topics that were controversial often included emotions like anger, afraid and annoyance. By using a regression model they concluded that positive results are more likely to be clicked.

The research done by Kim [20] looked at emotions from a user perspective. How does a user's emotion impact search behaviour? They conducted a user study with two search tasks. The results of their user study were that both the task and the perceived ability to control emotion have an effect on their behaviour [20]. General search tasks require more time and use more tools than specific tasks. The users with low emotion control and the perceived ability to control emotion [20], are less likely to handle pressure and are more inefficient during the search session. The researchers link this to the complex search task which can cause a burden on the users' emotions. This paper links the search behaviours with the emotional aspect of the user.

3

Well-Informed Users in the Wider World

To understand and define what a well-informed user is in the context of web search, we look beyond the domain of computer science. In other fields, there is already an attempt made to define a well-informed person. The context can range from medical aspects to more politically oriented topics. By comparing several of the definition and requirements needed in another field to be considered well-informed, we can synthesize a definition for our context of web search. The definition will not be very strict definition as that is impracticable to define. Nevertheless, we aim for a set of defining characteristics or features that a well-informed user should have.

The paper by Schutz [32] mentions what a well-informed citizen is. In his essay, he discussed some critical and necessary skills and attributes that a well-informed citizen should possess. These skills are required to deal with the magnitude of information. The description of being well-informed by Schutz [32] is having a reasonably founded opinion in whatever field they may be. There is a distinction made between the three types of individuals. First, an expert is someone whose judgements are based on assertions. On the opposite, we have the man on the street. Someone with knowledge in many fields, but they're not coherent. The knowledge is practical and enough for his needs. A well-informed citizen is someone in between the two types, possessing enough knowledge to have a reasonably founded opinion but not reaching the level of an expert. One of the features that a well-informed man possesses is the ability to recognize an expert and, more importantly, to come to a decision on his own after listening to an opposing expert. This suggests that there is enough knowledge to evaluate the information and compare it with their existing knowledge before deciding on an opinion. The construction of a well-informed individual is defined by knowledge, and the ability to leverage the knowledge to make decisions.

In the medical domain Marteau, Dormandy, and Michie [21] looked at measuring patients' informed choice. Different contexts, but the measurement aspect is something we need to do later in this thesis. We are interested in an informed attitude and how to know if it is informed. In a medical context, the authors approached it from the patient's choice. How can they know if the choice made by a patient is informed or not? They define informed choice as having all the information and alternatives used when deciding. Furthermore, this decision should be consistent with individual values. They need to understand the person's attitude and knowledge to measure this informed choice. The attitude in this specific context is towards a medical procedure. The rationale being a negative attitude means refusing a medical procedure if someone is informed. Therefore to measure informed choice, the authors measure three things knowledge, behaviour and attitude, which reflect the values. The resulting measurement was a binary categorization of knowledge. From the description of the measurements, knowledge is the most important. Attitude can play a role in knowledge retention and seeking, but that remains outside the measurements' scope.

In the previous chapter (Chapter 2, some used the definition of attitude while others used belief. Several papers discuss the definition of belief in a field outside of Computer Science. Particularly of interest is the definition of justified belief of interest in our search to define well-informed users. The

definition by Foley [12] says that to have a justified belief, it must full fill certain prepositions. To have a justified belief, the individual must have spent a reasonable amount of time and effort to come to this belief. This includes finding and evaluating other sources before coming to a justified belief. However, an exception allows one to claim a justified belief while spending little effort or time. This is only the case when the topic is of lesser importance. While the definition of justified belief is vague in regard to effort, time and evaluation, it does provide merit to the fact that knowledge is of the essence. Similar to the earlier definitions of informed individuals, knowledge is needed and used to evaluate information or evidence before coming to a conclusion.
Justified belief is applied in the context of web search by Miller and Record [23]. They note a few limitations to the requirement of time and effort; it should have an upper and lower bound to be considered justified. They explain that a web search user is responsible for knowing if the information is biased or incomplete. Even though users might be trapped in a filter bubble, they're still expected to filter the information before coming to a justified belief. Just because the information is not presented in the first search result is not a reason to claim a justified belief by only visiting the search result. To sum it up, in the context of web search, a justified belief requires the user to put in effort beyond the presented information and evaluate the results. This evaluation should look at bias and completeness. A belief formed by filtered search results risks the label of unjustified beliefs [23].

From our venture into the definitions of different fields, we find that the definition of a well-informed user has several features and attributes. The requirement of knowledge is very clear in all the different definitions. A certain level of knowledge is required in order to analyse and evaluate the received information. Furthermore, the individual should put in an effort to gather unbiased information. Using this as a requirement, we can expect a well-informed user to (1) have enough knowledge and (2) be knowledgeable and aware of the different biases regarding a topic. A way to measure this is to have a knowledge test or questionnaire. This is the clearest method for us to distinguish a well-informed user from an uninformed user. There could be other methods, such as behaviours, but from previous work, we know that detecting domain knowledge requires a longer time frame than a single search session to capture the behaviour [37]. In the later sections, we will look at the differences between well and uninformed users and reflect on the definition we decided to use for a well-informed user.

4 Experiments

4.1. Research process

The experiment consists of three parts, of which the first and second parts are preparations for the third part. The first experiment is to prepare the material needed to measure participants' level of informedness. It aims at **finding the right set of knowledge questions** that is used to differentiate the knowledge level of the participants. The second part consists of collecting the **annotation of search results** regarding their viewpoints. This information is required to decide which search results will be displayed on the search results page with and without a viewpoint-biased ranking. The third part aims at answering our research questions and consists of the **main user study** in which participants will do a search task and answer several questionnaires.

4.2. Preparation phase

4.2.1. Knowledge questions

This first experiment, question creation, aims to filter the questions that are most suitable to be used to categorize users. For the design of this experiment, a set of around 150 yes/no questions has been created on two topics. The topics are abortion and obesity. Each topic has between 60 and 80 questions. These two topics are the same topics that will be used in the main user study experiment. Both topics are sourced from ProCon[1], a webpage that lists controversial or debated topics.
Crowd-sourced participants will be asked to answer the dichotomous questions. The questions will include questions to test their attention and ensure that the answers are answered with the required level of attention. The participants will be given a set of 60 to 80 questions which takes less than 10 min to finish. Each participant is presented with questions on a single topic.

After the answers are collected, they are analysed to find the questions that tell us the most about the users' knowledge. For this, we will use methods such as Cronbach's α to find the questions that tell us the most about knowledge. Outliers, questions with a very high percentage of correct or wrong answers, will be removed from the set. The goal is to use the questions to separate the participants in the main experiment into three groups: uninformed, mildly informed and well-informed.

4.2.2. Viewpoints

To make the viewpoint-biased ranked search results page possible for the search task, we need to know which viewpoint the search results have on the debated topic. Each result must be carefully assessed on its stance towards the topic. The annotations are used to prefer some results over others depending on the experiment group during the search task. We collected the annotations from crowd workers. The search results, which are used for the annotation, are collected using Bing search API. Crowd workers can give an annotation based on a seven-point Likert scale. The options range from strongly opposing to strongly supporting a statement. The median of three viewpoint annotations per

[1] https://www.procon.org

result will be used. We will add an exclusion criterion if the majority of the responses are on both ends of the scale for a single search result.

4.3. Main user study

4.3.1. Materials

4.3.1.1 The questionnaires

The main experiment includes three questionnaires that will be asked twice to measure knowledge, emotion, and attitude. Two other questionnaires will be asked once before the search task. The first one, knowledge, has been obtained from the first experiment, questions creation. Read the earlier subsection 4.2.1 for more information.

SAM - self-assessment mannequin The second questionnaire is to ascertain the participant's emotional state before the start of the experiment. This experiment may confront some participants with information contrary to their beliefs and attitude on a topic. To understand the emotional impact of this experience, the participants are asked to answer questions relating to their emotional state using SAM, Self Assessment Manikin [3]. After the search task, the same questions will be asked again. The answer will be used to find if there is an emotional aspect to the change of beliefs. See figure 4.1 for the SAM.

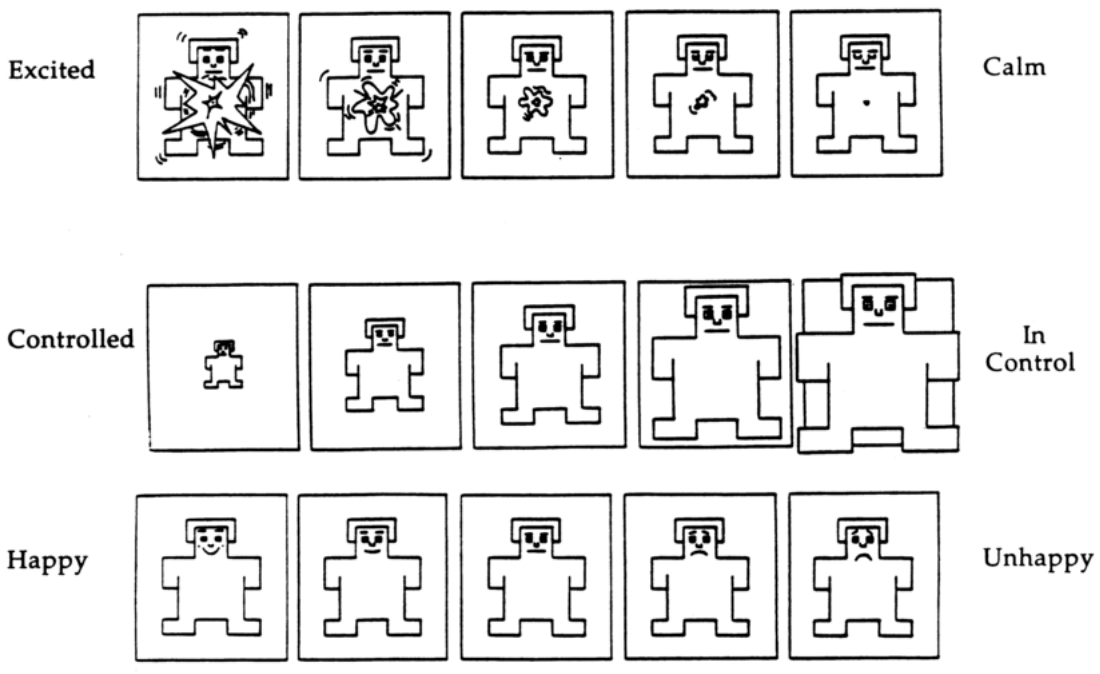


Figure 4.1: Self Assessment Manikin [3]

Attitude questionnaire The third questionnaire measures the participants' attitudes. To understand if there is any change in attitude due to the search task, the participants are asked about their attitude towards the two topics. The answers will be on a seven-point Likert scale. This question will be shown twice, before and after the search task. This makes it possible to detect attitude changes. The results will be used to answer the research question.

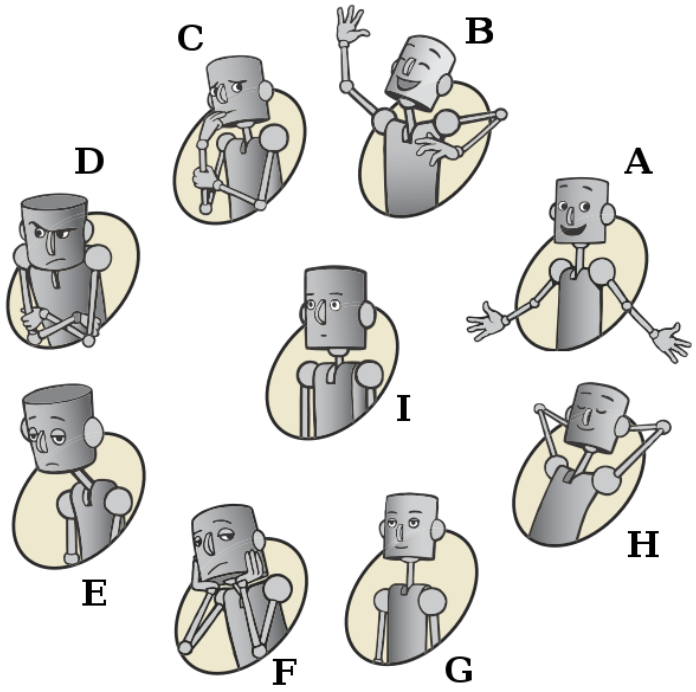


Figure 4.2: Pick-A-Mood pictograms with labels.

PAM and Receptiveness to opposing views Two other questionnaires will be included regarding mood and receptiveness to opposing views. For the mood questionnaire, we will make use of the PAM (pick a mood) questionnaire [5]. The participants choose from the pictograms in Figure 4.2 that describes their mood. The nine mood options are excited, cheerful, relaxed, calm, bored, sad, irritated, tense and neutral. The second questionnaire regarding receptiveness to opposing views contains 18 questions that help us understand if the participant is open to opposing views or not [25]. The results of both questionnaires will be analysed in an exploratory manner.

4.3.1.2 The Custom SERP

A custom SERP page has been created to control and track the participants' interaction with the different configurations of the SERP. It will be a static results page. The queries are predetermined and the results are annotated (not visible to the users) and processed during the preparation. Different viewpoint-biased ranked result pages will be shown depending on the participant's assigned experiment group. To track the interactions, we make use of the LogUI framework created by Maxwell and Hauff [22].

4.3.2. Procedure

Questionnaire The participants will be given all the questionnaires mentioned before the search task. The knowledge, emotion and attitude questions will be given again after the search task has been concluded. Some questions, like demographic, will not be asked directly since that information will be available from the crowd-working platform.

Search Task The participants will be provided with a search query relating to one of the two topics. The topics are abortion and obesity.

- Should Abortion Be Legal?
- Is Obesity a Disease?

During the main experiment, the questions in the questionnaire will be on a single topic. This is a random assignment to avoid bias in the selection of topics. The participants, independent of their answers

to the questionnaires, will be split into three types of search result groups. Depending on their group, the results of the search query will be manipulated to prefer certain viewpoints. The results will be biased or balanced regarding the search query. The first group will be shown a results page that is biased towards results that are in favour of the query. The second group will be shown results that do not favour the query. The third group will be shown an equal amount that is in favour and not in favour. For the biased results page, the ratio will be six-two-two. The first six will be biased to a viewpoint, the second two results will be neutral, and the final two will be opposing. The balanced view will consist of viewpoints split equally into two views.

The participants will be asked to collect arguments regarding one of the topics. For this, we will give the participant a scenario which requires the sourcing of arguments. The following scenario is provided to the participants:

"*You are participating in a mock debate with colleges. For this, you need to prepare arguments that you can use. Make use of the provided search results to find at least 3 arguments that you can use.*"

The search queries are predetermined, and the results are manipulated. Their collected arguments, and answers to the task, are not the experiment's focus. The focus is the effect on their attitude after the interactions with the search results. Their interaction will be tracked. This includes mouse movements, click behaviour and time spent on the SERP. We will log the interactions using the *LogUI* library by Maxwell and Hauff [22].

4.4. Experiment design

The main search experiment task is a factorial 3x3 between-subjects experiment. Each participant will only participate in one group. The variables and technical details of the experiment design are explained in this section.

4.4.1. Hypothesis

To answer the research question stated in Section 1.1, we test three different hypotheses. Depending on the result of the tests, we can answer our research question.

Hypothesis Attitude change

H1 : Users with different levels of informedness show different levels of attitude change in a web search.

H2a : Users who are exposed to results with a viewpoint-biased search ranking confirming their own prior bias show lower levels of attitude change than users that are exposed to unbiased (balanced) search results during a web search.

H2b : Users who are exposed to results with a viewpoint-biased search ranking conflicting with their own prior bias show higher levels of attitude change than users that are exposed to unbiased (balanced) search results during a web search.

H3 : The effect of biased compared to unbiased search result pages on attitude change is moderated by users' level of prior informedness.

4.4.2. Variables

4.4.2.1 Independent variables

In this experiment, we investigate the effect of two independent variables. Both are categorical.

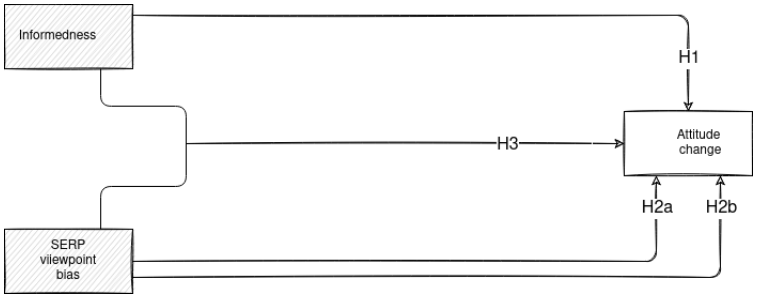


Figure 4.3: Graphical overview of the tested hypothesis. The grey boxes are the independent variables (IV) and the white boxes is the dependent variable (DV).

Table 4.1: Viewpoint ranking of search results

Result nr.	Viewpoint bias Supporting	Viewpoint bias Opposing	Viewpoint bias Balanced
1	Supporting	Opposing	Supporting
2	Supporting	Opposing	Opposing
3	Supporting	Opposing	Supporting
4	Supporting	Opposing	Opposing
5	Supporting	Opposing	Supporting
6	Supporting	Opposing	Opposing
7	Neutral	Neutral	Supporting
8	Neutral	Neutral	Opposing
9	Opposing	Supporting	Supporting
10	Opposing	Supporting	Opposing

Informedness level We will measure the informedness level of the participants. This variable is separated into three categories: uninformed, mildly informed, and well-informed. The allocation of participants is done through questionnaires. The participants are divided based on the percentile they score in. Those in the 70th percentile or above will be categorized as well-informed, the 30th to the 70th percentile as mildly informed, and all others as uninformed. Thus, this scale is relative to the other participants.

SERP Viewpoint Biased Ranking The second categorical independent variable is the search result page's bias. The search results page (SERP) for the experiment has been categorized into three groups. The SERP have an overall viewpoint biased towards supporting, opposing or balanced search results. Each viewpoint-biased SERP will have ten results. The biased viewpoints ranking is in order from top to bottom, 60% will be biased, 20% will be neutral, and the final 20% will be the opposite of the bias. The viewpoint bias towards balanced results will have equal supporting and opposing results in alternating order. See Table 4.1 for an overview.

4.4.2.2 Dependent variable

Attitude change After the search task, the participants will provide their attitude again. The difference in attitude or Δ attitude will be measured. This is a continuous result in percentages. The measurements will be done using a questionnaire where a single question will be asked per topic. The change prior to the search task and post the search task is Δ attitude.

4.4.2.3 Descriptive and exploratory variables

There are two variables measured to describe the population (age and gender) and several exploratory variables.

Age We obtain the ages of the participants through the crowd-sourcing platform. The goal is to understand the variation and spread of age groups. It allows us to understand whether our sample is representative of the general population.

Gender The second descriptive variable is the gender of the participants which is also obtained from the crowd-sourcing platform. The platform we use has two options male or female.

Knowledge gain As the interaction between the initially measured level of informedness and the SERP viewpoint bias, we want to know if it has affected the participants' knowledge. To measure the gain or loss in knowledge we will use our previously defined knowledge questionnaire. While the initial results for these questions will be used to categorize the participants, the results after the search task will be used to measure the gain in knowledge and do not affect the initial categorization of the participants. The score for knowledge gain (or loss) is a percentage.

Search interaction behavior

- Dwell time: Average time spent on the result page between actions
- Time spent: Total time spent on the search task
- nr Clicks: Measuring the number of clicks on results during the search task per participant.
- Mouse movements: Measuring the movements on the SERP (distance).

Emotion change To see the effect of the search interaction on the participant's emotional state, we will measure the differences in emotions. The changes in the emotional state of the participants will be assessed using SAM (Self Assessment Mannequin) [3]. The answers will be on a five-point Likert scale. The changes over the three measurements are represented by one value. Changes to more valence, dominance, and arousal are $+1$ while the other direction is -1 per step. The differences in emotion before and after the search task is the measurement taken for this variable.

Topic For each participant, we will also register which topic they've been assigned. This will be used to see if the participants show different responses based on the topic.

4.4.3. Participants

The participants will be sourced from Prolific[2]. Participating in the experiment is only possible after explicitly consenting. Participants are required to be fluent English speakers above 18 years of age. Each participant is allowed to participate in our study only once. For the main user study, we aim to balance the weak and strong prior attitudes. Our definition of strong attitudes is the four options on the ends of a seven-point Likert scale (two on both sides). The middle three options are our definition of weak attitudes. If our participant capacity has been reached for an attitude group (weak or strong), we will deny further participation.

4.4.3.1 Sample size

Sample size using power analysis tool G*Power [10] we find a required sample size of 315. The following options were used in G*power.

- Fixed effects, special, main effects and interactions
- Effect size Cohen's $f = 0.25$
- $\alpha = 0.05/4 = 0.0125$
- power = 0.8
- $Df = 4 = (informed\ levels - 1) \cdot (SERP\ viewpoint\ bias\ levels - 1)$
- groups = 9 = (3 levels of informedness)·(3 levels of SERP viewpoint biased ranking)

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

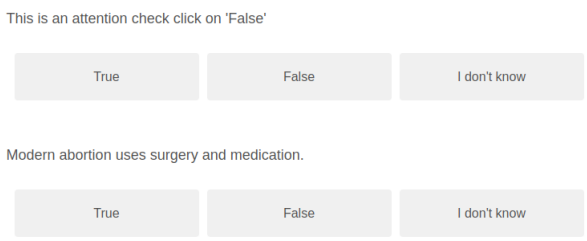


Figure 4.4: Knowledge question with an attention check

4.5. Implementation

4.5.1. Knowledge questionnaires

4.5.1.1 Questions

The choice was made to use dichotomous questions to construct the knowledge questionnaire. This setup of the questions allows phrasing questions or statements in such a way that there is only one possible answer. To avoid the possibility of guessing by the participants, there is an additional answering option included ”I don’t know”. All the questions are formulated in a similar matter, like a statement. The options to choose from are True, False, and I don’t know.

4.5.1.2 Survey

Per topic, twenty participants answered all the questions. In total 60 participants were sourced since there were initially three topics. There were three surveys, each on a different topic. Embedded within the roughly 70 questions there were attention checks. See Figure 4.4 for an example of the implementation. They were used to filter and reject participants that did not pay attention to the questions. The total duration for answering a single survey was around 6 minutes. Each participant was paid 9 GBP per hour.

4.5.1.3 Analysis

The internal reliability measurement was used to find a good subset of questions. Cronbach’s alpha method was used. Due to the computational limitation of being unable to create and calculate $C(n,r) = C(70,15) = 7.215 \cdot 10^{14}$ sets of questions, a random approach was used. Fifteen randomly chosen questions were picked and their Cronbach’s alpha was calculated. For each topic, 4000 options of fifteen questions were calculated. The final chosen set of questions all score above 0.8. This is a high score for internal consistency [34] and therefore these questions were chosen as the knowledge questionnaire. The list of questions can be found in Appendix B.

4.5.2. Viewpoint

4.5.2.1 Search results

The search results, to be annotated, were obtained from Bing search API [3]. For each of the three topics, two opposing search queries were used in order to obtain both supportive and opposing results. Per query, 15 search results were chosen to be annotated. Each result was checked manually to remove results that were hidden behind a paywall and login prompt. For results that contained minimal information, like one or two sentences were removed, as were duplicate results. The results list was checked from the top-ranked result to the lowest-ranked result. Finally, results for which an annotation was available were excluded from the annotation task [4]. All the results were obtained with the following queries.

- ”Is Human Activity Primarily Responsible for Global Climate Change?”

[3] https://www.microsoft.com/en-us/bing/apis/bing-web-search-api
[4] https://osf.io/v38c5

- ”Is Obesity a Disease?”
- ”Should Abortion Be Legal?”
- ”Is Natural Activity Primarily Responsible for Global Climate Change?”
- ”Is Obesity a Choice?”
- ”Should Abortion Be Illegal?”

4.5.2.2 Survey

Each survey consisted of 30 results that needed to be annotated. Each participant was asked to annotate 10 results to limit the risk of biased annotations. Each result required three annotations. A total of nine participants were required to annotate as a single topic, for the three topics, this amounts to 18 participants. Embedded with the annotation questions were attention checks formatted in a similar way as the search results. Each question had 7 options ranging from strong support to strongly oppose.

4.5.2.3 Analysis

There was an exclusion criterion used. If the participant failed the attention checks, their submission would be discarded. Furthermore, if a result had annotations that were on both extremes of the scale (Likert scale) then that result would be discarded as well. The last criterion did occur in one instance. The results obtained were given the median annotation out of the three annotations. Most of the annotated results were either positive/supportive or neutral in nature. Only the topic of abortion had several opposing results, which was enough to use for the custom SERP. The topic of obesity and climate change did not have many results with the annotation opposing. In fact, climate change had only one single result that was annotated opposing. This led to the conclusion that the topic of climate change was dropped from the user study experiment. While it is possible to find more opposing results, it is not the aim to find misleading results or disinformation, therefore it was dropped. The third topic, obesity, had 4 results that had opposing annotations. Fortunately, the results that were excluded were results with an opposing annotation. This meant there were enough results for both abortion and obesity to create and fill the viewpoint-biased ranked SERP.

4.5.3. Custom SERP

4.5.3.1 Architecture

To do the experiment a custom application was built. LogUI was integrated to provide the logging capabilities to capture the participant’s interaction with the SERP webpage. The web page was reachable at a custom sub-domain at sandboxedsearch.ewi.tudelft.nl. The application henceforth named sandboxed search, and LogUI were both hosted on the same virtual machine. To handle the routing between the two NGINX was configured as a reverse proxy. Sandboxed search was reachable under the route XXX/search and LogUI at XXX/logui. The reason for this choice was due to security implementation requirements. Some modern browsers do not let a user interact with a website unless it is an HTTPS connection. This is something the connection to the server required since the participants in the experiment can use any kind of browser. Fortunately, with the use of certbot[5] and its NGINX addon, the creation of SSL certificates turned out the be straightforward. It took several tries to get the right configuration for NGINX and required some tradeoffs, but once that was solved the connection to the server was secured. One of the tradeoffs was to embed the CSS and Javascript for the sandboxed search into its HTML-generated page. This has the added benefit that the number of requests to the server was minimized since everything was sent at once. Apart from the search page, the logging application needed its connection secured as well. It uses a WebSocket connection to log all the interactions. Again due to security measures from the browsers, connecting to a server using WS (websocket) from an HTTPS connection would be blocked. Therefore it had to be secured as well. Directly adding this to LogUI was one option, while the other was to use NGINX to secure the connection between the server and the user. The second option was implemented which meant that the connection to and from the server was WSS (websocket secure). Internally the connection from NGINX to the LogUI instance was

[5] https://certbot.eff.org/

an unsecured WS (websocket) connection. Since all three applications, NGINX, sandboxed search and Logui all resided on the same virtual machine the connection between them was not required to be secure.

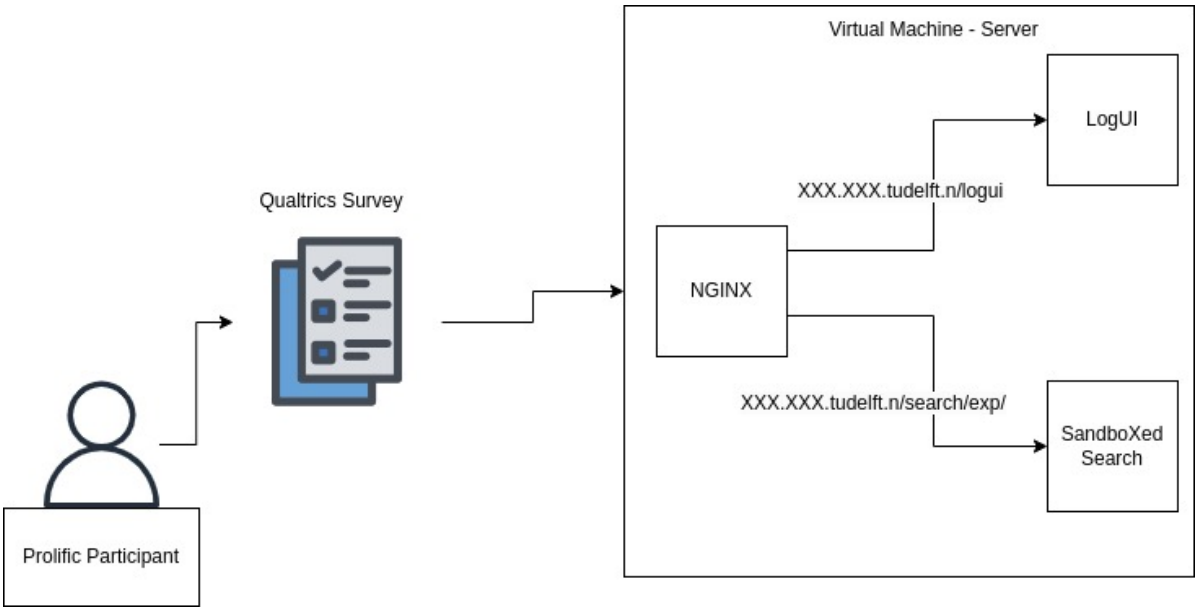


Figure 4.5: Overview of the components

4.5.3.2 SERP layout

Results The SERP was kept simple and uncomplicated. The participants would see a title, a search box and a list of results. Each SERP had ten results. Initially, the idea was to have more results on a second page, but that was removed due to insufficient annotated results. Hidden from the user in the HTML attributes, each result is annotated with its rank in the SERP and its viewpoint. The only visible aspect of a result is the title, link (complete URL) and a short summary (snippet). The title and link are clickable and open up in a new tab, this is to avoid stopping the logging process. See Figure 4.6 for a visualization.

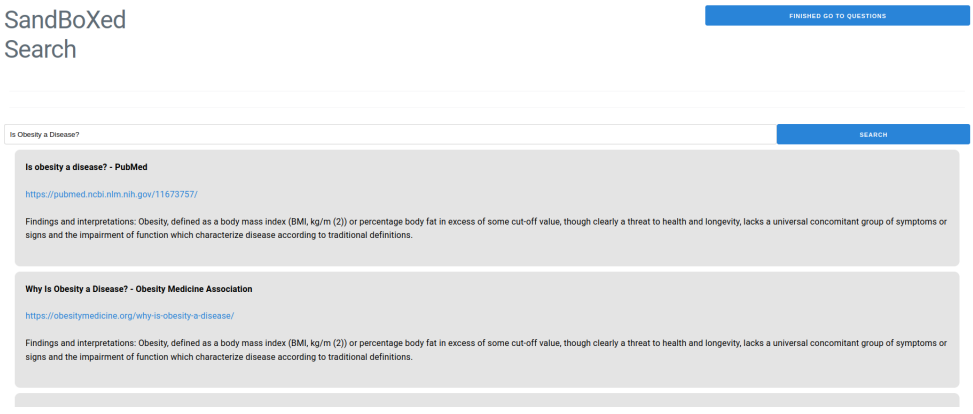


Figure 4.6: Custom SERP page upper part

Search option To give a convincing SERP, the search box and search button was included in the SERP, However, no action or behaviour was linked to those elements. In the search box, the statement

Figure 4.7: Pagination of the SERP

relating to the topic was shown. A participant could interact with it by altering the text, but nothing else would happen. The same applied to the search button.

Finished Once a participant was done, they could click the ”Finished Go back to questions” button. This would show an alert asking the user to close the tab. In turn, this would close the logging connection as well. The plan was not to show an alert, but to close the tab upon a click. Modern browsers do not allow the closing of a tab directly from Javascript unless it was opened using Javascript, which was not the case. Therefore the closing of the tab was requested to be done manually. The connection to LogUI would only close if the tab was closed as well, this is to avoid participants clicking ”Finished” too early and ending the logging early as well. Which would mean losing valuable interaction data. The risk is that a user leaves the page open long after finishing the survey.

Pagination At the bottom of the page, three buttons are added. These give the illusion of there possibly being another page. It was included to track users that were not pleased with the first ten results. To their surprise, they would only be scrolled back up to the top of the page. No new set of results becomes available. Figure 4.7 shows how it was presented to the participants.

4.5.3.3 URL parameters

The application is hosted and reachable through the domain name. Each variation of the SERP, be it topic or viewpoint bias is reachable on a separate page. To avoid making the users aware of the group they’ve been assigned, the URLs to SERP variations are shortened to two letters. It is not hidden from the participants that they are exposed to ’opinionated results’, but the viewpoint they’re assigned to is not explicitly publicised. For example, the SERP variation on the topic of obesity with a balanced (unbiased) viewpoint is denoted as OB, with the complete address becoming ”sandboxedsearch.XXX.XXXX.nl/search/exp/OB<PRAMETERS>”. The SERP accepts several parameters through the URL.

- $PROLIFIC\_PID$: This is the most important one as it relates to identifying the individual participants. It is needed to connect the user questionnaire answers to the users’ behaviour.
- $STUDY\_ID$: This one is the studyID. While not critically important, it does link back to which study the user participated in, In case later a new study is issued this will help separate the user from the other batches.
- $Option$: Here, all other data is put together that might be of interest during the analysis. It includes if the users’ attitude reply has been categorised as strong or weak. This is denoted by a 1 for strong and 2 for weak. The word Live, the SERP option and the sessionID is included as well. The SERP option is denoted by two letters.

The values of the parameters are passed along to the LogUI configuration. Thereby linking the interactions to the answers in the survey.

4.5.3.4 LogUI

LogUI [22] is used to track and log the interactions of the users. It consists of two components, the application hosted on a server and the client side of the application. The client side is embedded in the SERP HTML page and includes the logging configuration. Each aspect that needs to be logged must

be included in the configuration, otherwise, those actions of the user will not be logged. Critical is the configurations regarding the connections. There is no logging without those being correct. This includes the endpoint, where the server aspect of LogUI is hosted. As mentioned before it is a WebSocket connection to the application on the server. The following items were logged by LogUI;

- $Click\_RESULTS$: This would be logged when a participant clicked on a search result. The log includes the rank and viewpoint of the clicked search result.
- $Click\_RANDOM$: This would show in the log when the participants clicked somewhere in the SERP. From testing, this can include using the scroll wheel button. All clicks are recorded.
- $Hover\_IN/OUT$: Whenever a cursor hovers over a search result this event name would show up in the log. It includes which search result was hovered over, the rank and viewpoint.
- $Page\_click$: This logging is triggered when there is a click on the pagination buttons. For visualization see Figure 4.7

Apart from these specific mouse events, LogUI registers the cursor position every 100ms. This event logs the X and Y coordinates of the cursor. Another important piece of information logged is when the page is in focus and when it's not. Whenever the SERP page is left to visit another tab, it will record it. This can be used to approximate how long each result has been visited. However, due to the nature of the task, which includes the writing or typing of the answer, it is unreliable to use it to measure the duration of visiting a result. Measuring the time spent on the SERP page is reliable since it cannot be mistaken for another action.

4.5.4. Crowd study

4.5.4.1 Ehtical considerations

This experiment includes human participants. A requirement for such studies is to have the experiment approved by Human Research Ethics Committee at the TU Delft. This includes which data is collected, how, and what is done with the data. The participants must be presented with the option to participate or not after knowing what is collected and for what reason. Before the participants start the experiment, they're asked to give consent. Only after explicit consent can they participate. They are made aware of the topic in advance and that they will be presented with 'opinionated' results. After the experiment, they are given a debrief. It is explicitly mentioned that the results were manipulated to show a viewpoint. They are given another option to revoke their participation in the experiment or continue. If they choose to opt-out their collected data will be removed. The consent statement can be found in Appendix A

4.5.4.2 Strong and Weak attitudes

The participants are split into two groups before participating. During a prescreening study, the participants are asked to give their attitudes towards the two topics and statements. Those responses are used to balance the pool of participants over the topics. This allows for a group of participants with different attitudes to participate. This does play a role in the randomization of the SERP viewpoint bias. Randomization is done within a group of strong attitudes or weak attitudes. Thereby spreading both types of attitudes equally over the different SERPs.

5

Results

The total number of participants in the main user study was 320. Of which 5 submissions were invalid due to technical error when directing them to the SERP, which was caught during the smaller pilot deployment with 20 participants per topic. Another set of 24 submissions was excluded since they did not visit the SERP, this was not due to technical issues. The total of complete submissions is $320 - 5 - 24 = 291$.

5.1. Descriptive variables

However, combining all the interactions with the demographic data we obtain 287 submissions with data on their age. This means that there are 4 submissions that are not part of the sample population description below.
The options for gender were male, female and preferred not to say. However, in some cases, the demographic data included "$DATA\_EXPIRED$" as gender. Both the "data expired" and "prefer not to say" are grouped under other in the table 5.1. The descriptive variables are further split by SERP condition in table 5.2. The descriptive variables shown are on the 287 submissions with information on their age. The distribution of the age of the participants see Figure 5.1, the distribution separated per topic can be found in Figure 5.2

Table 5.1: Descriptive Statistics Age and Gender

	Age			Participants			
	Median	Mean	Std	Total count	Male	Female	Other
Abortion	26.0	28.49	8.04	140	57.857%	41.429%	0.714%
Obesity	25.0	28.09	8.18	147	48.299%	49.660%	2,041%
Combined	25.0	28.28	8.10	287	52.962%	45.645%	1.394%

Furthermore, we can break down the age and gender data further by the assigned SERP viewpoint bias.

Table 5.2: Descriptive Statistics Age and Gender per SERP viewpoint-bias

	Age			Participants			
	Median	Mean	Std	Total count	Male	Female	Other
Abortion Support	25.5	28.36	7.95	50	62.0%	38.0%	0%
Abortion Opposed	27.0	28.15	6.953	47	57.4%	40.4%	2.128%
Abortion Balanced	25.0	29.0	9.317	43	53.488%	46.51%	0%
Obesity Support	25.0	28.4468	8.5460	47	51.063%	46.808%	2.128%
Obesity Opposed	26.0	28.816	7.707	49	48.980%	51.0204%	0%
Obesity Balanced	24.0	27.0588	8.34125	51	45.098%	50.980%	3.922%

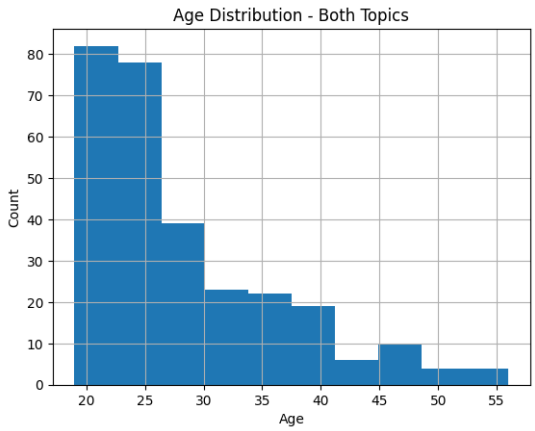


Figure 5.1: The age distribution over both topics combined

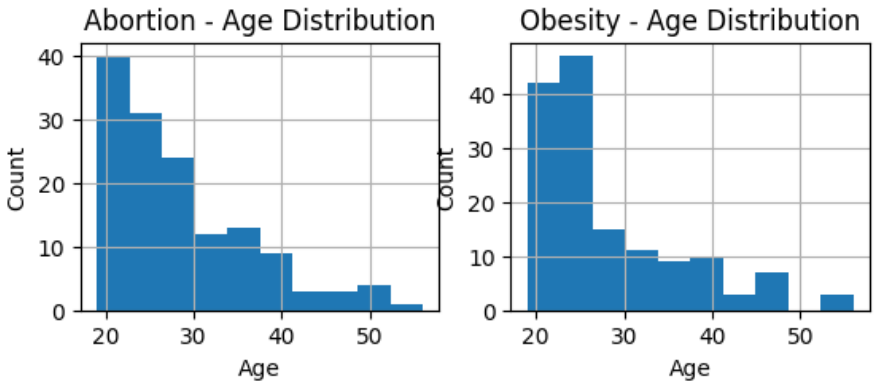


Figure 5.2: The age distribution over both topics separated

5.2. Attitude changes

Attitude change is measured as the difference between the initial attitude and the attitude after the search task interaction. The value ranges from -6 to +3, where positive values mean a strengthening of the initial attitude. Vice versa, the negative values mean a weakening of the initial attitude. The attitude changes can be analysed by topic and informedness levels. Before the statistical test can be done, we need to ensure that we can take the entire dataset as one set. Therefore we look at the attitude changes by topic. The means and standard deviation of both topics should be similar, otherwise, we need to analyse them as two separate groups. In Figure 5.4 we can see that the means are very similar. There is a small difference in standard deviations. The exact numbers can be found in Table 5.3. In Figure 5.3 we visualize the distribution of attitude change over the topics. Figure 5.5 compares the means of attitude change per informedness and has the topics separated as well. The average attitude change is 0.203 (SD=0.933) for well-informed participants, -0.0342 (SD=1.131) for mildly informed participants and -0.227 (SD=1.041) for uninformed participants. In total 108 participants changed their attitude, which is 37%. The change resulted in a stronger attitude for 64 and a weakened attitude for 44 participants. The remaining 183 participants did not change their attitude.
There were 118 participants with a 'weak' attitude. See Subsection 4.4.3 for the explanation of 'strong' and 'weak' attitudes. The other 173 participants had a 'strong' attitude. Of those 173 participants, 23% changed their attitude and 77% did not change their attitude. However, of the participants with a weak attitude, 58% changed their attitude while 42% did not. Of the well-informed participants, 55 had a strong attitude while only 14 changed their attitude. There were 24 who had a weak attitude, of which 15 changed their attitude. Of the 32 uninformed participants with a strong attitude 6 changed attitudes, from the 34 uninformed participants that had weak attitudes nearly half 16 showed an attitude change. There were 86 mildly informed participants that had a strong attitude, of which 20 changed

their attitude. From the 60 mildly informed participants with a weak attitude, an attitude change was registered by 37 participants.

	Abortion	Obesity	Combined
Mean	0.0	-0.027	-0.014
SD	0.772	1.291	1.071

Table 5.3: Attitude changes compared over topics

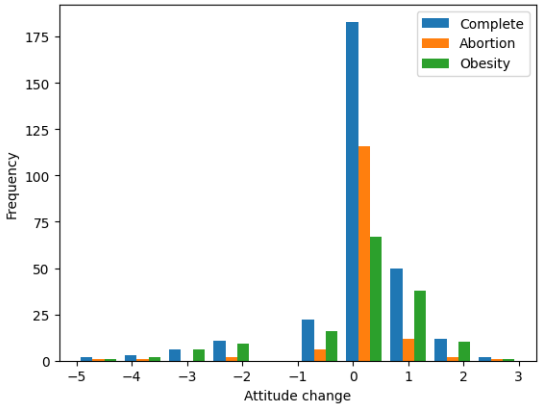


Figure 5.3: Distribution of attitude change over topics

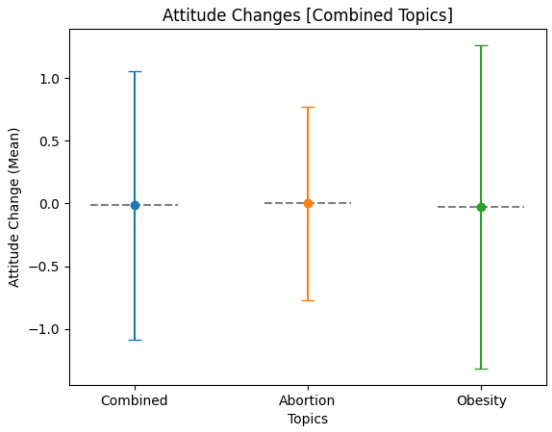
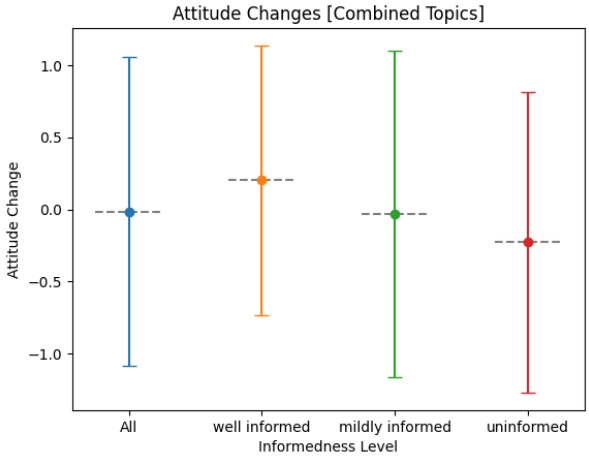


Figure 5.4: Means and standard deviation per topic

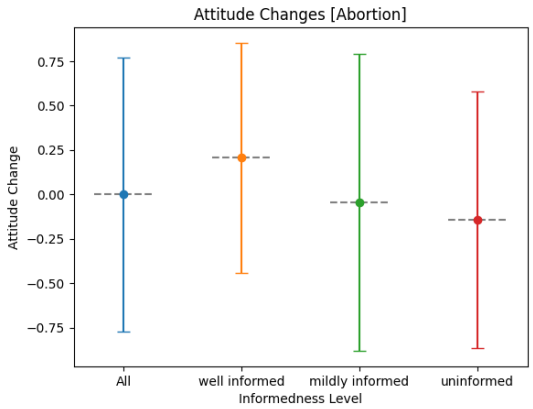
5.3. SERP viewpoint bias

One of the two independent variables in this experiment was the search viewpoint bias. Each participant was assigned to one randomly. In Figure 5.6 the differences in attitudes are compared by SERP viewpoint bias. The viewpoint bias is with regards to the topic statement in Section 4.3.2. To also compare if there is a topical influence, the users per viewpoint bias are separated into the assigned topics. Table 5.4 shows the values for the mean and SD per combination of topic and viewpoint bias. Just like the attitude change per informedness level, the attitude change is corrected for pre-existing bias. The participants with a neutral bias (value 0) are included and any change in their attitude is seen as a strengthening since they have no prior bias.

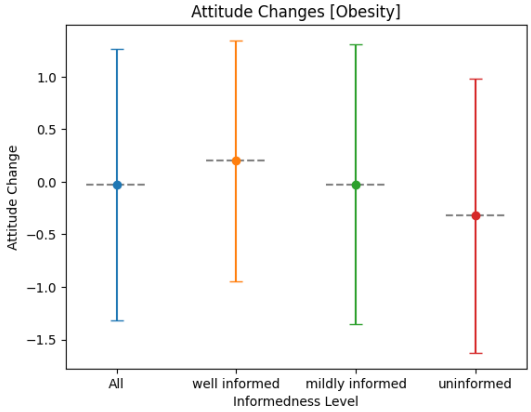
The attitudes of the participants per viewpoint bias are nearly similar. For the SERP viewpoint-bias support, 79% of the participants had a supportive attitude, 12% had an opposing attitude and 9% had a neutral attitude. For the SERP viewpoint-bias opposing, 77% of the participants had a supportive



(a) Attitude change by informedness



(b) Attitude change by informedness for abortion



(c) Attitude change by informedness for obesity

Figure 5.5: Attitude change by informedness levels

attitude, 17% had an opposing attitude and 6% had a neutral attitude. For the SERP viewpoint-bias balance, 82% of the participants had a supportive attitude, 12% had an opposing attitude and 6% had a neutral attitude. The attitudes of the participant and the SERP viewpoint bias are regarding the statements of the topic mentioned in Section 4.3.2.

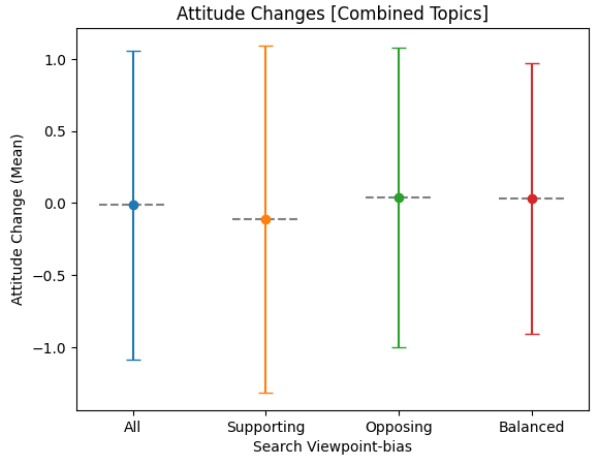
	Abortion	Obesity	Combined
Support	Mean : 0.04 SD : 0.774	Mean : -0.265 SD : 1.509	Mean : -0.111 SD : 1.205
Opposed	Mean : 0.064 SD : 0.665	Mean : 0.020 SD : 1.301	Mean : 0.04167 SD : 1.0400
Balanced	Mean : -0.113 SD : 0.859	Mean : 0.154 SD : 0.988	Mean : 0.031 SD : 0.940

Table 5.4: Attitude change per search-viewpoint bias

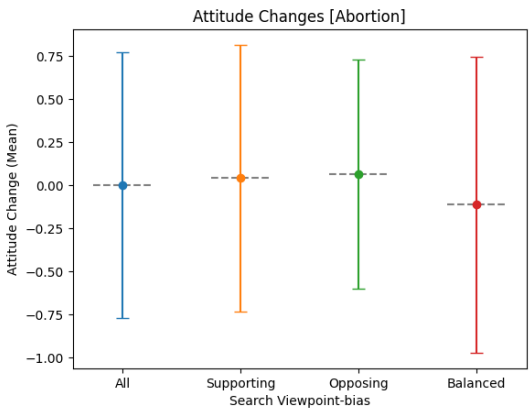
5.4. Hypothesis testing

To test our hypotheses, we will conduct a two-way ANOVA for attitude change. In case of significant observations with a Bonferroni-Holm correction applied [16] ($\alpha = 0.0125$), we will conduct a post hoc test for further analysis. This is only for H1 and H3. For H2a and H2b, we will use a different test to compare the two groups. This an addition that was not included in the preregistration of the experiment[1].

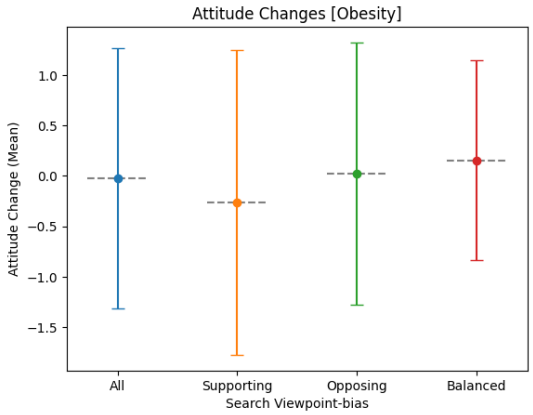
[1] see https://osf.io/g3r5z for the registration prior to the experiment



(a) Attitude change by viewpoint-bias



(b) Attitude change by viewpoint-bias for abortion



(c) Attitude change by viewpoint-bias for obesity

Figure 5.6: Attitude change by viewpoint-bias

5.4.1. H1 and H3

To analyse and understand the gathered data, we used visual materials and calculated statistics such as the mean and standard deviation. See the Figures 5.4, 5.5 and 5.6. The first hypothesis H1 and the third hypothesis H3, are tested using a 2-way ANOVA. In table 5.6, we can see the results for the 2-way ANOVA. The independent variable ”SearchViewpointBias” will be addressed by H2a and H2b. The two-way ANOVA shows that there is not a significant difference by (H1) informedness on attitude change (F = 3.02, p = 0.05). The p-value must be lower than the alpha value ($\alpha = 0.0125$) to have significant results. From the ANOVA results, we also find that the interaction effect (H3) of informedness and search viewpoint bias does not have a significant effect on attitude change (F = 1.00, p = 0.41). In Figure 5.7, we plotted an estimation plot. The group with mildly informed participants was used as a control group in the estimation plot. The attitude changes have a similar shape, and the group of well-informed users do reach higher. However, the mean difference shows a small overlap between the two which can explain the effects of informedness we have observed.

5.4.2. H2a and H2b

To test the two hypotheses, we do not consider the submission of the participants with an initial attitude of 0. This reduces the submission to 270 from 291. The participants are separated into groups based on their prior bias in relation to the assigned search viewpoint bias. Those assigned the balanced SERP viewpoint bias will be our unbiased group mentioned in hypotheses H2a and H2b. H2a(b): Users who are exposed to results with a viewpoint-biased search ranking confirming (opposing) their own prior bias show lower (higher) levels of attitude change than users that are exposed to unbiased or balanced search results during a web search.

Nr.	Hypothesis
H1	Users with different levels of informedness show different levels of attitude change in a web search
H3	The effect of biased compared to unbiased search result pages on attitude change is moderated by users' level of prior informedness

Table 5.5: Hypothesis H1 and H3

	sum_sq	df	F	PR(>F)
C(Informed_level)	6.883	2.0	3.023	0.050
C(SearchViewpointBias)	1.545	2.0	0.678	0.508
C(SearchViewpointBias):C(Informed_level)	4.560	4.0	1.001	0.407
Residual	321.074	282.0	NaN	NaN

Table 5.6: Anova results table

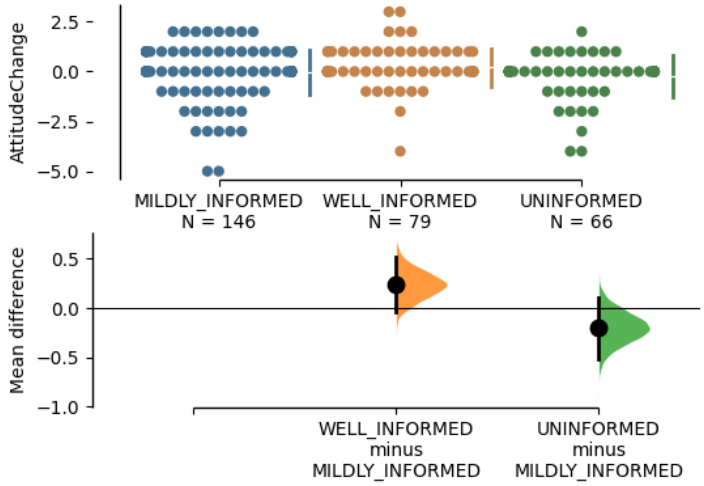


Figure 5.7: Estimation plot of attitude change

SERP bias type	Mean and standard deviation
Supporting bias	Mean :-0.021 SD:1.005
Opposing bias	Mean :-0.011 SD:0.942
Balanced	Mean :-0.244 SD:1.168

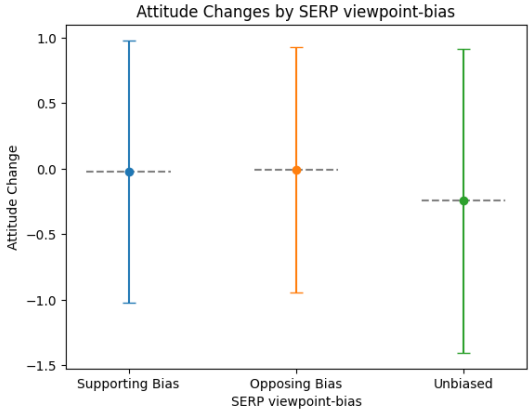
Table 5.7: Mean and deviation per SERP bias configuration

For the statistical testing, we cannot use the t-test as the assumptions of normality and equal variances are not met. Therefore we test H2a and H2b with a non-parametric test Kruskal Wallis. It showed no significant results (Statistic: 2.272 p-value: 0.132) for H2a. The same insignificant results are also observed in the results for H2b (Statistic: 1.635 p-value: 0.201). For a better understanding, the attitude changes per bias towards the participant have been visualized in Figure 5.8

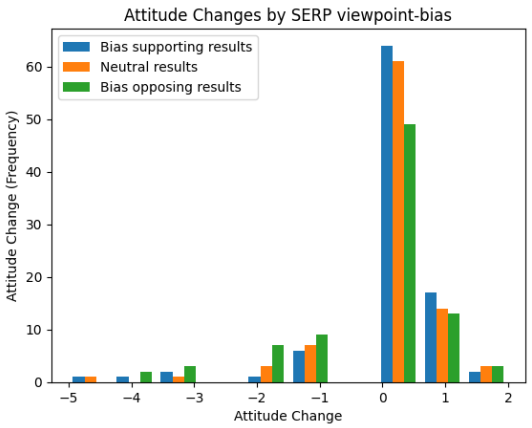
5.5. Exploratory variables

5.5.1. SERP mousemovements

In Table 5.9, the means and SD are listed. The distance is calculated by taking the sum of the distances between the x and y coordinates recorded in the logs. The differences between the different informedness groups are not large, especially taking into account the standard deviation. For a more visual comparison, see Figure 5.9, where the y-axis is the frequency. Most participants are labelled mildly informed, which explains the higher frequency counts. Larger differences can be found when



(a) Attitude change by viewpoint-bias towards participant



(b) Distribution of attitude change per viewpoint-bias towards participant

Figure 5.8: Attitude change by bias towards participant

the mouse movements of different SERP pages are compared. In Table 5.9, the mean and standard deviation of each SERP are presented. In table 5.8, we split the participants into those that changed their attitude and those that did not, this includes participants that had an initial attitude of '0' (neutral).

Attitude	Mean	SD
Changed	23986.335	16875.381
Unchanged	19835.214	17569.981

Table 5.8: Mousemovements of attitude changed and unchanged participants

Informedness	Mean	SD (standard deviation)
Well Informed	21095.304	17068.644
Mildly Informed	21903.667	17056.896
Uninformed	20512.226	18728.301

Table 5.9: Distance mouse movements by informedness

SERP	Mean	SD (standard deviation)
Abortion Support	17975.508	12528.236
Abortion Opposed	20629.906	18283.058
Abortion Balanced	19808.023	20388.795
Obesity Support	23785.838	20751.250
Obesity Opposed	23276.115	17801.961
Obesity Balanced	20788.345	13949.591

Table 5.10: Distance mouse movements by informedness

5.5.2. SERP time

The time spent on the SERP is logged and can be categorized by informedness level. Some participants never closed the SERP, meaning there is no end time. Those participants are left out of the analysis on time. Furthermore, there are participants that could have left their SERP open long after they were done. This does explain the outliers who spent a very long time on SERP, such as 245 and 202 minutes. The calculation for the mean and standard deviation is only done on the values within the 95% confidence interval. The well-informed participants spent, on average 9.97 minutes (SD = 6.90). This is

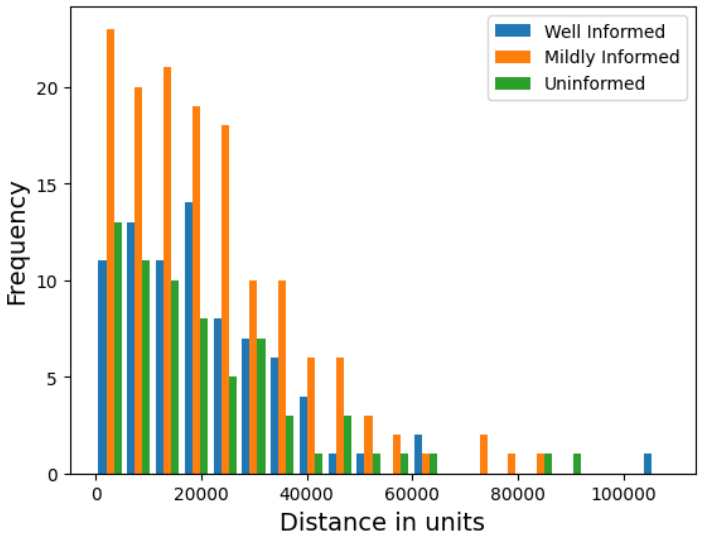


Figure 5.9: Mouse movement distribution per informedness level

longer than the mildly informed participants, who spent on average 9.54 minutes (SD= 7.51). The least time spent on the SERP belongs to the third group of uninformed participants, with an average time of 8.89 (SD=7.03). However, the differences are nearly within a minute of each other. The distribution of the time spent per informed group can be found in Figure 5.10.
The time spent can also be analysed from the perspective of a topic. The average time spent on the SERP with the topic ”Abortion” was 9.020 minutes (SD=6.794). For the topic ”Obesity”, the average time was a little bit more at 10.405 minutes (SD=9.388). When we compare the time spent by the participants that changed their attitude and those that did not, we find similar values. The participants that changed spent an average of 10.544 minutes (SD= 8.951), while those that did not have an attitude change spent an average of 9.046 minutes (SD= 6.854).

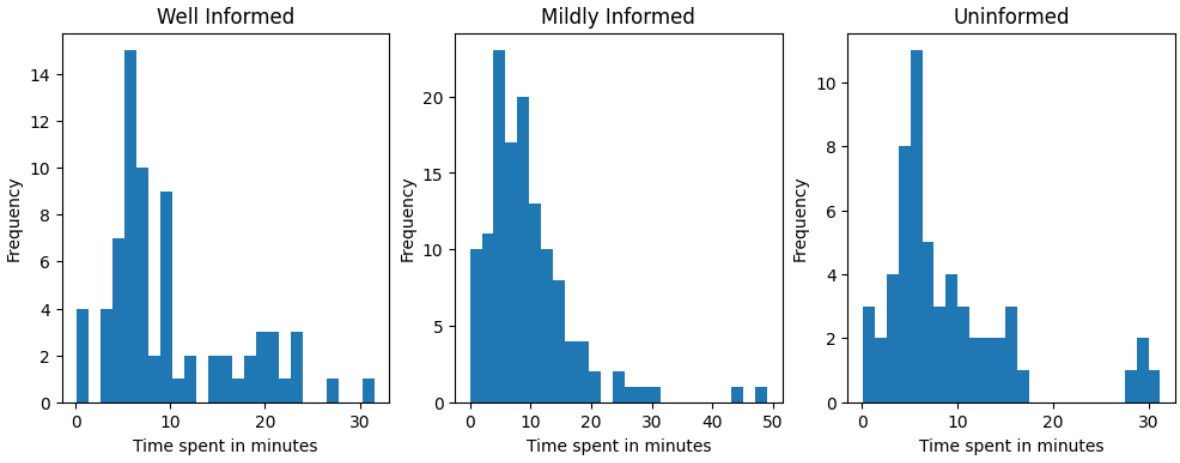


Figure 5.10: Time spent per informed group

5.5.3. SERP clicks

Figure 5.11 below shows the number of clicks on a result per SERP page. Each rank has a certain viewpoint. See Table 4.1 for the overview per SERP viewpoint bias. The spikes in Figure 5.11a correspond with results with a supporting viewpoint. The increase in clicks in Figure 5.11c at the results 8 and 9 (on the SERP, these would be the last two results) also have a supporting viewpoint. The peak at the sixth result is one with a neutral viewpoint. The comparison in the number of clicks on the results is

listed in table 5.11, the results are based on the values within the 95% confidence interval. There are, however, few differences in click interaction when the participants are categorized per informedness group. The mildly informed participant clicked on average on 4.1 results (SD=3.3), the well-informed participants on 4.7 results (SD=3.6), and the uninformed participants clicked on 4.4 results (SD=2.9). All of the groups had a median number of 4 clicked results.
The clicks were logged with their viewpoint annotation towards the topic statements. It creates the option to see the viewpoint clicked per stance of the participants in relation to the prior attitude on the assigned SERP. Participants with neutral attitudes (attitude score 0) are excluded from this part of the analysis. The participants that were assigned a SERP viewpoint-bias that supported their stance clicked 81% (M = 3.15, SD= 2.56) of all clicks on results that supported their attitude, 8% (M = 0.33, SD= 0.58) that opposed their attitude and 11% (M = 0.43, SD= 0.62) on neutral results. Those assigned to SERP viewpoint-bias that oppose their stance clicked 27% (M = 1.08, SD= 1.17) of all clicks on results that supported their attitude, 52% (M = 2.16, SD= 2.16) that opposed their attitude and 22% (M = 0.88, SD= 0.92) on neutral results. for the SERP Abortion support is 3.6 (SD=4.0),

Topic and Bias	Median	Mean	SD (standard deviation)
Abortion Support	2	3.0	2.9
Abortion Opposed	3	3.4	2.2
Abortion Balanced	4	4.7	2.8
Obesity Support	4.5	5.4	4.2
Obesity Opposed	4.5	4.8	3.2
Obesity Balanced	4	4.0	2.8

Table 5.11: Amount of clicks per SERP viewpoint-bias

5.5.4. Knowledge gain

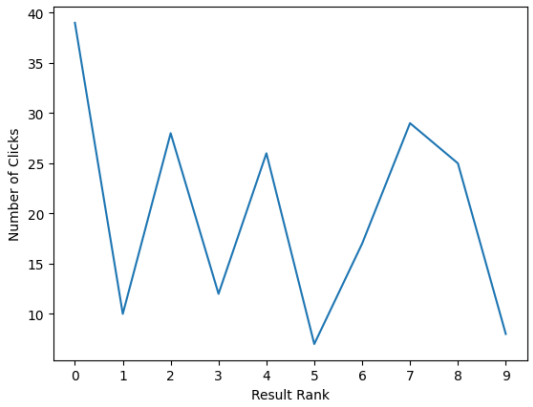
The participants filled in a knowledge questionnaire before and after the search task. The difference between the two is the knowledge gain that we measure and report. For the topic of abortion, there was an average knowledge gain of -0.007 (SD=1.788). The search pages with the topic obesity had an average knowledge gain of 0.713 (SD=1.659). The knowledge gain per informedness level can be found in 5.12. In table 5.13, the knowledge gain per search viewpoint option is shown. The knowledge gain or losses per combination of informedness and search viewpoint-bias is shown in table 5.14. Knowledge gain for those who changed attitudes was, on average, 0.403 (SD=1.774). For those that remained with their initial attitude, the knowledge gain was, on average, 0.318 (SD=1.742). Overall, 41.2% of the participants showed a knowledge gain, 31.6% showed a knowledge loss, and 27.1 % showed neither a knowledge loss nor gain. For participants with a 'weak' attitude, 47.5% gained in knowledge, 29.7% had knowledge loss and 22.9%. It was more evenly distributed among the participants with strong attitudes. A knowledge gain was shown by 37.0%, a loss in knowledge by 32.9% and no change in knowledge by 30.1% of the participants with a strong attitude.

Informed level	Mean	SD (standard deviation)
Mildly informed	0.486	1.749
Well informed	-0.380	1.215
Uninformed	0.985	2.011

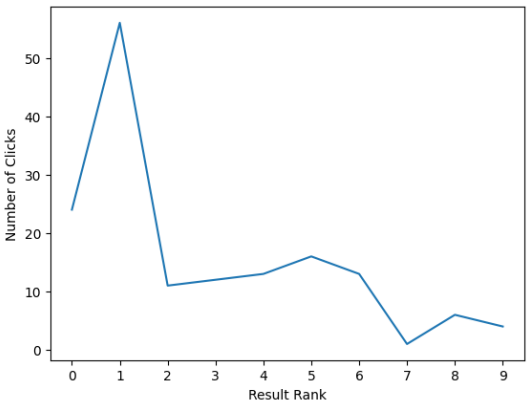
Table 5.12: Knowledge gain/loss per informedness level

Viewpoint-bias	Mean	SD (standard deviation)
Opposing	0.116	1.653
Supporting	0.769	1.864
Balanced	0.238	1.699

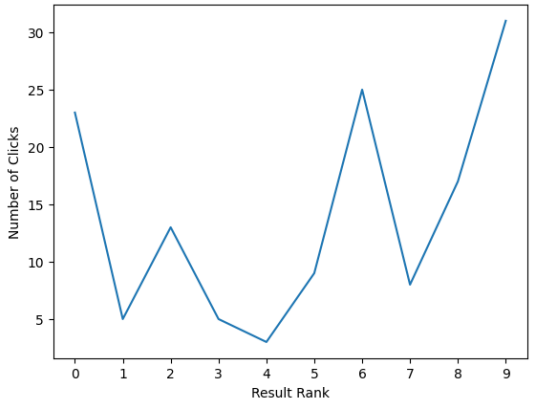
Table 5.13: Knowledge gain/loss per search viewpoint-bias



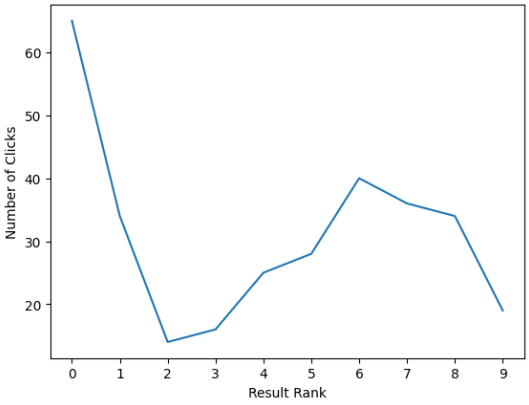
(a) Number of clicks per rank for the SERP Abortion Balanced



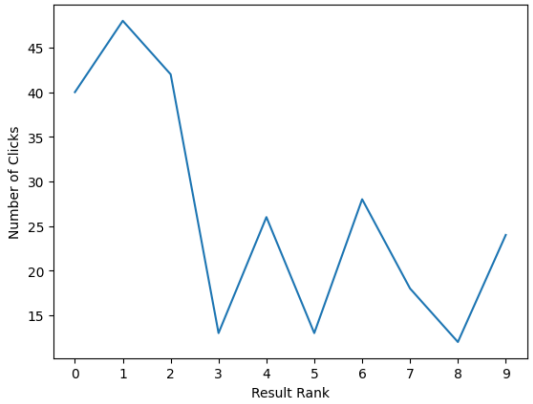
(b) Number of clicks per rank for the SERP Abortion Support



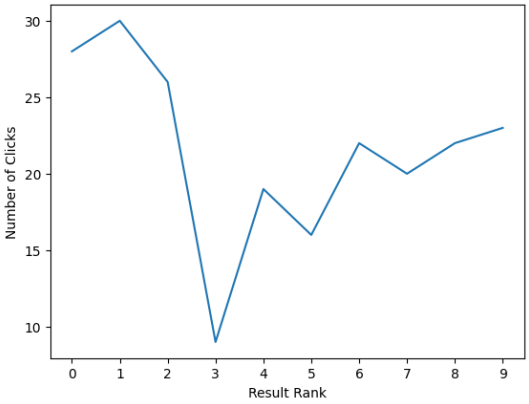
(c) Number of clicks per rank for the SERP Abortion Opposed



(d) Number of clicks per rank for the SERP Obesity Balanced



(e) Number of clicks per rank for the SERP Obesity Support



(f) Number of clicks per rank for the SERP Obesity Opposed

Figure 5.11: Number of clicks per SERP viewpoint-bias separated by topics

5.5.5. Emotion

For each participant, there were three questions on emotion. They were on valence, arousal and dominance. In table 5.15, you can find the changes in emotions over the different informedness shown. A negative value means a decrease in valence, arousal or dominance. More details regarding emotional changes per SERP can be found in Table 5.16. See Table 5.17 for the changes per topic. In Table 5.18 the changes in emotions are listed for those with and without attitude change.

5.5.6. Receptiveness to opposing views

In the papers by Minson and Chen [24] and Minson, Chen, and Tinsley [25], they use a scale to measure how open someone is to opposing views. In the user study, it was presented to the participants as a

Informdness level	SERP viewpoint-bias	mean	SD (standard deviation)
Mildly informed	Abortion Supporting	0.042	1.485
Mildly informed	Abortion Opposed	0.238	1.998
Mildly informed	Abortion Balanced	0.045	1.637
Mildly informed	Obesity Opposed	0.385	1.841
Mildly informed	Obesity Supporting	1.130	1.895
Mildly informed	Obesity Balanced	0.933	1.340
Well informed	Abortion Supporting	-0.25	1.362
Well informed	Abortion Opposed	-0.765	1.352
Well informed	Abortion Balanced	-0.8	0.872
Well informed	Obesity Opposed	-0.1	1.136
Well informed	Obesity Supporting	-0.118	1.078
Well informed	Obesity Balanced	-0.231	1.120
Uninformed	Abortion Supporting	0.643	1.950
Uninformed	Abortion Opposed	1.0	1.826
Uninformed	Abortion Balanced	-0.167	2.544
Uninformed	Obesity Opposed	1.462	1.599
Uninformed	Obesity Supporting	2.0	1.563
Uninformed	Obesity Balanced	1.333	1.491

Table 5.14: Knowledge gain per informedness for each SERP option.

Informedness level	Type Emotion	Mean	SD (standard deviation)
Mildly informed	Valence	-0.120	0.642
Mildly informed	Arousal	0.034	0.725
Mildly informed	Dominance	0.062	0.821
Well informed	Valence	-0.139	0.689
Well informed	Arousal	0.025	0.811
Well informed	Dominance	0.076	0.792
Uninformed	Valence	-0.152	0.557
Uninformed	Arousal	0.091	0.690
Uninformed	Dominance	0.076	0.840

Table 5.15: Type of emotion change per informedness level

Search viewpoint-bias	Type Emotion	Mean	SD (standard deviation)
Supporting	Valence	-0.071	0.537
Supporting	Arousal	0.0	0.841
Supporting	Dominance	0.040	0.665
Opposing	Valence	-0.052	0.698
Opposing	Arousal	0.083	0.672
Opposing	Dominance	0.146	0.989
Balanced	Valence	-0.260	0.649
Balanced	Arousal	0.052	0.698
Balanced	Dominance	0.021	0.763

Table 5.16: SAM changes per SERP viewpoint-bias

questionnaire. Taking the 95% confidence interval of the results we got the mean value per group. For those that changed their attitude, their average score was 3.019 (SD= 12.436), and the participants that did not change their attitude had an average score of 2.663 (SD=12.922). The averages show more difference when grouped by informedness. Well-informed participants had an average score of 3.961 (SD=15.140), and mildly informed participants had an average score of 1.883 (SD=11.872). Participants in the category uninformed had a mean score of 2.889 (SD=11.807).

Topic	Type Emotion	Mean	SD (standard deviation)
Abortion	Valence	-0.163	0.615
Abortion	Arousal	0.092	0.693
Abortion	Dominance	0.043	0.807
Obesity	Valence	-0.093	0.657
Obesity	Arousal	0.0	0.783
Obesity	Dominance	0.093	0.827

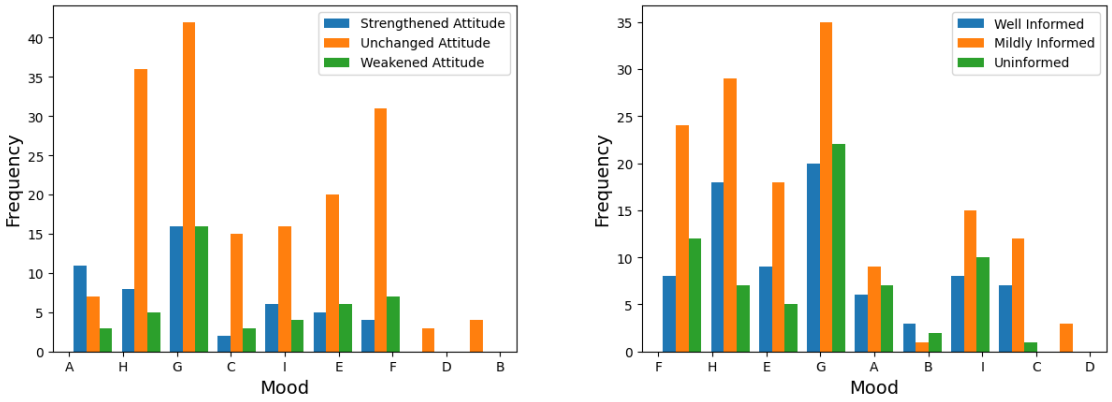
Table 5.17: Changes in SAM [3] per topic

Attitude Change	Type Emotion	Mean	SD (standard deviation)
Changed Attitude	Valence	-0.130	0.682
Changed Attitude	Arousal	0.037	0.719
Changed Attitude	Dominance	0.056	0.731
Unchanged Attitude	Valence	-0.126	0.610
Unchanged Attitude	Arousal	0.049	0.756
Unchanged Attitude	Dominance	0.077	0.865

Table 5.18: The changes in emotion with and without attitude change

5.5.7. Mood

At the start of the experiment, the participants were asked to provide their mood using PAM [5]. The options for mood are shown in Figure 4.2. The distribution per informedness level is shown in Figure 5.12b, and the distribution per attitude change situation can be seen in Figure 5.12a. In the distribution where prior bias is considered, such as Figure 5.12a, neutral attitudes are excluded.



(a) Mood choices per attitude change

(b) Mood choices per informedness category

Figure 5.12: Distribution of mood per attitude change and informedness category

6

Discussion

6.1. Findings and implications

We investigated the literature to formulate the definition of a well-informed user. To understand the effect of informedness on attitude change we conducted a user study in which we compared participants' attitudes before and after the search task. The search task required participants to list three arguments using the provided SERP. In the provided SERP, the results that were presented were manipulated to be biased towards a viewpoint. This was reflected in the ranking of the results as well. Interaction with the SERP was logged per participant. Each participant also answered several questionnaires before and after the search task.

RQ 1 From the literature, we found at least two requirements that should be met by a user to be considered well-informed. Both requirements are on a single topic.

1. Enough knowledge
2. Be knowledgeable and aware of the different biases or arguments

These requirements are limited in practicality since they require input from the user to be determined. This means that every topic needs its own specialized set of questions to see if users meet the criteria of well-informed users. During the experiment, the participants were measured against each other, this is another constraint in the application of the measurement. Therefore there is a necessity to look into other characteristics that define well-informed users, otherwise, the usage of the definition may stay limited to scoped experimental settings.
For the applicability of the definition, we examine the exploratory findings to see if there is a difference in behaviour per informedness level that can be observed. Based on the literature, we know that differences in time spent and interaction can be observed between domain experts and novices [37]. However, looking at the results we obtained no clear differences in the behaviours of the participants when categorized by informedness. There was a similar distribution in mouse movements, even the means showed strong similarities. For the distribution, see 5.9. Time spent on the SERP pages was also similar across the different informedness groups. The means were all close to each other. From the logged click actions, we found that the number of clicks on the results was similar. The average clicks for well-informed participants were slightly higher than the other groups at 4.7. The differences are, however, too small to make meaningful conclusions. Each group had an average of between 4 and 5 clicks. As anticipated, the knowledge gain was, on average, the highest for uninformed participants. A knowledge loss for well-informed users was unexpected. While a knowledge gain was expected to be minimal, a loss was not foreseen. This could be either due to the search results that were shown and thus had an adverse effect or the failure of the knowledge test to capture their knowledge accurately. That being said, seeing an average loss while the questions remained the same is surprising. On the emotional aspect, the changes are very similar. All three informedness levels, on average, decreased in valence and increased slightly in arousal and dominance.
No difference between the different informedness levels could be found based on the SERP interaction

and questionnaires collected during this research. This is surprising since we expected to see more noticeable differences in interaction behaviour. Users with domain knowledge have shown different behaviours than those without in previous studies [17]. Therefore, there was an expectation of seeing the same. This could be due to the limited interaction the custom SERP provided. Participants were not allowed to construct a query and were given a predetermined results list.
Overall, the initial definition of well-informed users seems to hold up. On average, this group of participants had a slightly higher score on the scale of receptiveness to openness, which can be seen as a weak acknowledgement of the second requirement. This does need to be further researched by explicitly asking questions that relate to the different arguments given a topic. This was not done in detail during the experiment to avoid priming the participants. However, each topic included a question relating to arguments that other viewpoints could use. We avoided asking more detailed questions on arguments. Otherwise, it would not be clear whether the change was due to a search interaction or the reflection and discussion on a topic. Furthermore, we did look to see if there was any behavioural difference in the SERP, as we expected based on previous research, but we found little to no difference in SERP interaction. More freedom on the SERP and query formulation might show this, but that was outside of the focus of this experiment.

RQ 2 With the second research question, we aim to find differences in attitude changes. Hypotheses H1, H2a, H2b and H3 are to support answering the research question. The first hypothesis (H1) tested whether any effect is measurable in attitude changes by informedness level. As shown in the results, there was no significant measurable effect. In the estimation plot in Figure 5.7, we can see that there is a very small overlap between the two plotted mean differences.

Hypotheses H2a and H2b looked at the prior attitude in regard to the SERP viewpoint bias. That, too, did not show any significant results. In hindsight, looking at the click distribution per rank, it became clear that participants sought the desired results. Even though the first results did have a high click rate, in some cases the lower-ranked results had a higher click rate (see Figure 5.11c). Research has shown that order ranking can affect users [2], which did not happen in this case. During the search session, they were tasked to find at least three arguments. It wasn't mentioned which stance they should argue for. This was by choice since the goal was to have as natural SERP interaction as possible with the limitations in place. We observed that participants looked for results matching their prior attitudes. This was especially the case when participants in the experiment were shown a SERP with a similar viewpoint bias as their initial attitude. Over 80% of all clicks were on results that aligned with their attitude.

Hypothesis H3 revealed no significant interaction effect between informedness and SERP viewpoint bias. This means that the combination of the two independent variables did not significantly affect the participants' attitude change.

We need more conclusive results to answer the second research question on the susceptibility of well-informed users to attitude compared to other informedness levels. More than a third of the participants changing their attitude should be explainable, yet not one of our hypotheses showed a significant result. There needs to be additional research using another experiment design to exclude the effect of informedness on attitude change. For example, in a study, participants are tasked with a search task on multiple topics for which they have varying levels of informedness. This would allow comparisons across topics per informedness and a single participant over different informedness levels. If the results are consistent with our results, this would have a significant impact on SERP results and how search engines deal with presenting the information.
We know from our interaction with search engines that SERPs have been improved by presenting information cards to give short, unbiased information on some topics [31]. The implications of these results could mean that providing more information, thereby increasing knowledge, might not be the way to support users in dealing with bad or misinformation. Furthermore, since our results suggest that informedness does not seem to influence attitude change, other strategies might be better suited to support users in dealing with conflicting information. One such strategy is providing a comparison [42]. This is one of the options mentioned in works that looked at increasing critical information seeking, which was covered in Chapter 2.

6.1.1. Other findings

Figure 5.11 shows that some SERPs show distinct click behaviour. This is especially visible on the topic of Abortion, with the SERP viewpoint bias balanced and opposed. In both cases, the ranks with supportive results get more clicks than others. This results in the peaks that are observable in Figure 5.11a. This is tied to the attitudes of the participants, the majority were supportive of the topic statement. For the viewpoint-bias balance, most participants (82%) had a supportive attitude towards the topic statement. During the search task, they were tasked to find at least 3 arguments. However, since it was never mentioned which stance they should argue for, it could explain the increased activity with results that have a similar viewpoint, even if they were ranked lower.
We found differences between the informedness categories regarding the mood and the scale for receptiveness to opposing views. A higher score suggests more willingness to interact with opposing views [25]. The largest difference was between well and mildly informed. This could hint that mildly informed have some existing knowledge determining their attitude. In contrast, the average scale of the category uninformed is higher than mildly informed. The suspicion is that they might be more willing to engage with opposing views due to their limited knowledge of the topic. However, to confirm this suspicion, more research needs to be done. The difference between those that changed and those that did not change their attitude was very small on the scale.
According to the literature, most attitude changes should have occurred with participants with a weak attitude [36]. We did observe this as well within our results. This is in line with our expectations. Even though this suggests that attitude change is more prevalent in participants with a weaker attitude, it does not address the cause of attitude change. Our results do not seem to indicate that there is a connection between informedness and weak attitude. Regardless of informedness, the participants with weak attitudes changed more often in attitude compared to participants with strong attitudes. Over all three informedness categories, roughly half of the participants with a weak attitude had an attitude change.

6.2. Limitations

There were several limitations during the experiment. For one, there were only two topics, and both were related to the medical field. The attitude changes could have been influenced by having prior experience or relation to the topic. Someone who is obese might respond differently than someone who has no intersection with obesity. Using topics outside the medical domain can avoid or minimise the influence of prior associations. For example, other topics like deep sea mining might affect everyone differently, but it does not excessively affect a single individual. This is not the case with abortion and obesity. Certain individuals are affected, while others are not.
Furthermore, the topics were controversial. By design, we aimed for topics participants knew of, albeit at varying levels. Other topics that do not include controversial topics would be ideal to support the generalisation of the results. This would make the conclusions and results more widely applicable than currently is the case.
The SERP the participants interacted with was severely limited, partially due to the annotation requirement in the experiment setup. That, too, can be improved by having annotations be more than binary options as suggested by [6]. More annotated results could support more results and queries. This allows for more interaction and, in turn, a more interactive search session. Another aspect is to allow free query formulations in the SERP. The query lengths and the number of formulations in that setup can show differences. Most likely, the behavioural differences would be more pronounced between different levels of informed users.
Another limitation is that most participants were supportive towards the topic statement. To cover all options, it would be prudent to have a balanced group and consider the prior bias when looking at attitude change. This was not taken into consideration during the experiment setup, as the focus was on informedness and not on prior attitudes.

Finally, a more robust knowledge questionnaire could prove to be valuable. While the current setup did work as expected, we have not tested for different amounts of questions. It remains a risk that there were only 15 questions, where a minor change in one or two questions could have a larger effect than is needed. Another concern is that the questions were not tailored to the results. This means that not all of the knowledge gains could be captured. The more basic questions on the topic and possible

solutions or options were covered by some results. However, the more detailed questions, such as historical trends, were unlikely to be answered using the provided search results.

7

Future Work and Conclusion

7.1. Future work

While this work aimed to understand the role of informedness in attitude change, there is a need for more work to complete our understanding. We found that our results suggest that informedness does not influence attitude change. Future work is needed to corroborate our findings or show that this is indeed the case. One aspect that can be improved upon is topics outside the medical domain. Both topics were related to the medical domain. This could have impacted our experiment if some of the participants had prior relation to one of the two topics while others did not. Furthermore, since the topics used were controversial, other topics that are not or less controversial might give a better insight. Another aspect is to use a more interactive SERP. As is already stated in Section 6.2, the current setup of the SERP did not allow for query formulations since we controlled the SERP.
Even though our definition of well-informed users did hold up, a more rounded definition should include the behavioural aspect. More interactions with the SERP could lead to finding differences in behaviour. This might eventually remove the current dependency on knowledge tests. More work could be done to automatically find or estimate the user's knowledge level and informedness. This can be used to support participants by suggesting queries or results that can be used to increase their knowledge.

7.2. Conlcusion

In this thesis, we examined the role of informedness in attitude change during web searches. Through a user study, our results suggest that informedness does not influence attitude change. This has implications for the way search engines support their users in dealing with controversial topics. Just providing unbiased knowledge may not be enough to support users. An option would be to show different biases as a comparison or overview as well. However, we still do not know what causes attitude change during a web search. Therefore this still remains an unresolved gap in the current research into attitude change in web search. More work is needed to look further into informedness and attitude change by, for example, considering a range of topics beyond the medical domain that are not or less controversial than the one used in the experiment.
In the search task during the experiment, each participant was assigned a viewpoint-biased ranking. While influenced by the ranking, the exploratory results showed that participants eventually followed their prior bias and clicked on their desired result even if they were ranked at the bottom of the SERP.

This thesis also aimed to define a well-informed user in the context of web search. We are able to conclude that the two criteria we deduced based on literature from other fields do apply to the context of web searches. A well-informed user on a topic should be knowledgeable and understand the different arguments on that topic. Future work can look further into this and define behavioural characteristics as well. This should make the use of the definition possible outside of experimental conditions.

A

Consent

A.1. Consent form

The following statement was presented to participants in the main user study. A similar approach was used for the participants in the viewpoint and knowledge surveys. Their consent statement differed in description and data collection.

CONSENT FORM

You are being invited to participate in a research study titled " How is attitude change influenced by users informedness in web search". This study is being done by master student Suleiman Kulane from the TU Delft. The purpose of this user study is to understand the role of knowledge (being informed) on attitude change and will take you approximately 10-15 minutes to complete. The data will be used for a Master's thesis and possible future scientific publications.

We will log the search session and ask you to answer questions relating to your emotions, existing knowledge and attitude (opinion) on one of two topics. The topics are abortion and obesity. After interacting with possibly opinionated search results on a topic, you will be asked once again on knowledge, emotion and attitude (opinion).

As with any online activity, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. We will minimize any risks by altering the IDs into randomly generated ID. The original IDs will be removed after the user study has been concluded. There will be no link back to the participants. Some of the data, such as age and gender, will only be published as aggregated data. The data will be published, after identifiable information has been removed, on the TU delft repository.

Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to stop at any point during the user study. This will mean a withdrawal from the user study due to incomplete answers. The withdrawal means that data gathered during the user study will be removed. Because the participants are anonymous we are not able to treat individual requests. For any additional questions, you can contact xxxxx@student.tudelft.nl.

By clicking continue you agree to participate in the user study which includes the processing and publication of the results. If you do not consent, please press disagree to exit.

A.2. Debrieving

Please read this carefully

The opening statement mentioned opinionated results. It was not clarified what was meant with opinionated results. For clarification, the search results were manipulated to randomly be balanced or favour one viewpoint over another. We hope that withholding this clarification on search manipulation did not cause any discomfort or unease. If this clarification has made you reconsider your participation, then you can do so by clicking on ”withdraw participation and consent”. The data gathered will be deleted and your participation in the experiment will be cancelled. Otherwise press “continue” to submit and finish the experiment.

Thank you for your participation and time.

B

Knowlege items

Below the knowledge items for topics of abortion and obesity are listed. Attention checks are not included. Those were automatically generated when the questions were converted into a template using a simple script.

No.	Question
1	”Of the abortions done worldwide every year, 45% is done unsafely.”
2	”Over a third of the 205 million pregnancies each year are unintended.”
3	”Around 56 million abortions are performed each year around the world.”
4	”The methods used for abortion are depended on legality, availability and doctor or woman’s preference.”
5	”In the United States the maternal mortality is higher after abortion than after childbirth.”
6	”The method used to abort a pregnancy depends on the stage of the pregnancy.”
7	”Reasons for late terminations of pregnancy, after 20 weeks, include birth defects and risks to the woman’s health.”
8	”The pro-choice movement emphasizes a woman’s right to bodily autonomy.”
9	”The vast majority of miscarriages occurs before medical practitioners can detect an embryo.”
10	”Abortion is the termination of a pregnancy by removal of an embryo or fetus.”
11	”Stillbirths and premature birth are generally considered miscarriages.”
12	”Modern abortion uses only medication.”
13	”Spontaneous abortion can be caused by accidental trauma.”
14	”On average abortion is safer than carrying a pregnancy to term.”
15	”In the United States the maternal mortality is more than 10 times lower after abortion than after childbirth.”

Table B.1: Abortions questions

No.	Question
1	”Obesity can only be solved using a medication or surgical approach.”
2	”Those with obesity have a higher chance of cancer.”
3	”Obesity only occurs in low-income countries.”
4	”Obesity more common in men than women.”
5	”Fixing obesity is solely a medical issue.”
6	”Obesity is linked to ethnicity.”,
7	”In the US more than 1 in 3 adults are considered to have obesity.”
8	”Obesity is something to strive towards.”,
9	”Obesity is another name for being fat.”,
10	”Higher educated groups have a higher obesity prevalence.”
11	”Obesity rates have declined by half between 1980 and 2015.”
12	”Old people cannot have or get obesity.”
13	”Diet is a treatment for obesity.”
14	”Obesity can be prevented.”
15	”Obesity is not classified as a disease.”

Table B.2: Obesity questions

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”. In: *Journal of Medical Internet Research* 16.4 (Apr. 2014), e2642. DOI: 10.2196/jmir.2642.

[2] Leif Azzopardi. “Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval”. In: *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 2021, pp. 27–37.

[3] Margaret M. Bradley and Peter J. Lang. “Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential”. In: *Journal of Behavior Therapy and Experimental Psychiatry* 25.1 (Mar. 1994), pp. 49–59. ISSN: 00057916. DOI: 10.1016/0005-7916(94)90063-9.

[4] Jack Conklin. *A taxonomy for learning, teaching, and assessing: A revision of Bloom’s taxonomy of educational objectives complete edition*. 2005.

[5] Pieter MA Desmet, Martijn H. Vastenburg, and Natalia Romero. “Mood Measurement with Pick-A-Mood: Review of Current Methods and Design of a Pictorial Self-Report Scale”. In: *Journal of Design Research* 14.3 (2016), pp. 241–279.

[6] Tim Draws et al. “Comprehensive Viewpoint Representations for a Deeper Understanding of User Interactions With Debated Topics”. In: *ACM SIGIR Conference on Human Information Interaction and Retrieval*. CHIIR ’22. New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 135–145. ISBN: 978-1-4503-9186-3. DOI: 10.1145/3498366.3505812.

[7] Tim Draws et al. “This Is Not What We Ordered: Exploring Why Biased Search Result Rankings Affect User Attitudes on Debated Topics”. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Virtual Event Canada: ACM, July 2021, pp. 295–305. ISBN: 978-1-4503-8037-9. DOI: 10.1145/3404835.3462851.

[8] Dima El Zein and Célia da Costa Pereira. “User’s Knowledge and Information Needs in Information Retrieval Evaluation”. In: *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. Barcelona Spain: ACM, July 2022, pp. 170–178. ISBN: 978-1-4503-9207-5. DOI: 10.1145/3503252.3531325.

[9] Robert Epstein and Ronald E. Robertson. “The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections”. In: *Proceedings of the National Academy of Sciences* 112.33 (Aug. 2015), E4512–E4521. DOI: 10.1073/pnas.1419828112.

[10] Franz Faul et al. “Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses”. In: *Behavior research methods* 41.4 (2009), pp. 1149–1160.

[11] Leon Festinger. *A Theory of Cognitive Dissonance*. Stanford University Press, 1957. ISBN: 978-0-8047-0911-8.

[12] Richard Foley. *Justified Belief as Responsible Belief*. na, 2005.

[13] Ujwal Gadiraju et al. “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. New York, NY, USA: Association for Computing Machinery, Mar. 2018, pp. 2–11. ISBN: 978-1-4503-4925-3. DOI: 10.1145/3176349.3176381.

[14] Souvick Ghosh, Manasa Rath, and Chirag Shah. “Searching as Learning: Exploring Search Behavior and Learning Outcomes in Learning-Related Tasks”. In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. 2018, pp. 22–31.

[15] Eddie Harmon-Jones. “A Cognitive Dissonance Theory Perspective on the Role of Emotion in the Maintenance and Change of Beliefs and Attitudes”. In: *Emotions and Belief: How Feelings Influence Thoughts*. Studies in Emotion and Social Interaction. New York, NY, US: Cambridge University Press, 2000, pp. 185–211. ISBN: 978-0-521-77138-2 978-0-521-78734-5. DOI: 10.1017/CBO9780511659904.008.

[16] Sture Holm. “A simple sequentially rejective multiple test procedure”. In: *Scandinavian journal of statistics* (1979), pp. 65–70.

[17] Christoph Hölscher and Gerhard Strube. “Web Search Behavior of Internet Experts and Newbies”. In: *Computer networks* 33.1-6 (2000), pp. 337–346.

[18] Samuel Ieong et al. “Domain Bias in Web Search”. In: *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*. WSDM ’12. New York, NY, USA: Association for Computing Machinery, 2012, pp. 413–422. ISBN: 978-1-4503-0747-5. DOI: 10.1145/2124295.2124345.

[19] Gabriella Kazai, Paul Thomas, and Nick Craswell. “The Emotion Profile of Web Search”. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR’19. New York, NY, USA: Association for Computing Machinery, July 2019, pp. 1097–1100. ISBN: 978-1-4503-6172-9. DOI: 10.1145/3331184.3331314.

[20] Kyung-Sun Kim. “Effects of Emotion Control and Task on Web Searching Behavior”. In: *Information Processing & Management*. Evaluation of Interactive Information Retrieval Systems 44.1 (Jan. 2008), pp. 373–385. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2006.11.008.

[21] Theresa M. Marteau, Elizabeth Dormandy, and Susan Michie. “A Measure of Informed Choice”. In: *Health Expectations* 4.2 (2001), pp. 99–108. ISSN: 1369-7625. DOI: 10.1046/j.1369-6513.2001.00140.x.

[22] David Maxwell and Claudia Hauff. “LogUI: Contemporary Logging Infrastructure for Web-Based Experiments”. In: *European Conference on Information Retrieval*. Springer, 2021, pp. 525–530.

[23] Boaz Miller and Isaac Record. “Justified Belief in a Digital Age: On the Epistemic Implications of Secret Internet Technologies”. In: *Episteme* 10.2 (2013), pp. 117–134.

[24] Julia A. Minson and Frances S. Chen. “Receptiveness to Opposing Views: Conceptualization and Integrative Review”. In: *Personality and Social Psychology Review* 26.2 (May 2022), pp. 93–111. ISSN: 1088-8683. DOI: 10.1177/10888683211061037.

[25] Julia A. Minson, Frances S. Chen, and Catherine H. Tinsley. “Why Won’t You Listen to Me? Measuring Receptiveness to Opposing Views”. In: *Management Science* 66.7 (July 2020), pp. 3069–3094. ISSN: 0025-1909. DOI: 10.1287/mnsc.2019.3362.

[26] Alamir Novin and Eric Meyers. “Making Sense of Conflicting Science Information: Exploring Bias in the Search Engine Result Page”. In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. CHIIR ’17. New York, NY, USA: Association for Computing Machinery, Mar. 2017, pp. 175–184. ISBN: 978-1-4503-4677-1. DOI: 10.1145/3020165.3020185.

[27] Rik Peels. “Responsible Belief: A Theory in Ethics and Epistemology, by Rik Peels, New York, Oxford University Press, 2017”. In: *International Journal of Philosophical Studies* 26.4 (Aug. 2018), pp. 601–643. ISSN: 0967-2559. DOI: 10.1080/09672559.2018.1511146.

[28] Frances A. Pogacar et al. “The Positive and Negative Influence of Search Results on People’s Decisions about the Efficacy of Medical Treatments”. In: *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. ICTIR ’17. New York, NY, USA: Association for Computing Machinery, Oct. 2017, pp. 209–216. ISBN: 978-1-4503-4490-6. DOI: 10.1145/3121050.3121074.

[29] Suppanut Pothirattanachaikul et al. “Analyzing the Effects of Document’s Opinion and Credibility on Search Behaviors and Belief Dynamics”. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. CIKM ’19. New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 1653–1662. ISBN: 978-1-4503-6976-3. DOI: 10.1145/3357384.3357886.

[30] Rod D. Roscoe et al. “Online Information Search and Decision Making: Effects of Web Search Stance”. In: *Computers in Human Behavior* 56 (Mar. 2016), pp. 103–118. ISSN: 0747-5632. DOI: 10.1016/j.chb.2015.11.028.

[31] Sara Salimzadeh, David Maxwell, and Claudia Hauff. “The Impact of Entity Cards on Learning-Oriented Search Tasks”. In: *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*. 2021, pp. 63–72.

[32] Alfred Schutz. “The Well-Informed Citizen”. In: *Collected Papers II*. Vol. 15. Dordrecht: Springer Netherlands, 1976, pp. 120–134. ISBN: 978-94-010-1342-0 978-94-010-1340-6. DOI: 10.1007/978-94-010-1340-6_6.

[33] Masaki Suzuki and Yusuke Yamamoto. “Analysis of Relationship between Confirmation Bias and Web Search Behavior”. In: *Proceedings of the 22nd International Conference on Information Integration and Web-Based Applications & Services*. iiWAS ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 184–191. ISBN: 978-1-4503-8922-8. DOI: 10.1145/3428757.3429086.

[34] Mohsen Tavakol and Reg Dennick. “Making sense of Cronbach’s alpha”. In: *International journal of medical education* 2 (2011), p. 53.

[35] Ryen White. “Beliefs and Biases in Web Search”. In: *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2013, pp. 3–12.

[36] Ryen W. White. “Belief Dynamics in Web Search”. In: *Journal of the Association for Information Science and Technology* 65.11 (2014), pp. 2165–2178. ISSN: 2330-1643. DOI: 10.1002/asi.23128.

[37] Ryen W. White, Susan T. Dumais, and Jaime Teevan. “Characterizing the Influence of Domain Expertise on Web Search Behavior”. In: *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. WSDM ’09. New York, NY, USA: Association for Computing Machinery, Feb. 2009, pp. 132–141. ISBN: 978-1-60558-390-7. DOI: 10.1145/1498759.1498819.

[38] 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*. New York, NY, USA: Association for Computing Machinery, July 2020, pp. 91–100. ISBN: 978-1-4503-7098-1.

[39] 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*. New York, NY, USA: Association for Computing Machinery, June 2021, pp. 240–244. ISBN: 978-1-4503-8366-0.

[40] Takehiro Yamamoto, Yusuke Yamamoto, and Sumio Fujita. “Exploring People’s Attitudes and Behaviors Toward Careful Information Seeking in Web Search”. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. CIKM ’18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 963–972. ISBN: 978-1-4503-6014-2. DOI: 10.1145/3269206.3271799.

[41] Yusuke Yamamoto and Satoshi Shimada. “Can Disputed Topic Suggestion Enhance User Consideration of Information Credibility in Web Search?” In: *Proceedings of the 27th ACM Conference on Hypertext and Social Media*. 2016, pp. 169–177.

[42] Yusuke Yamamoto and Takehiro Yamamoto. “Query Priming for Promoting Critical Thinking in Web Search”. In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. CHIIR ’18. New York, NY, USA: Association for Computing Machinery, Mar. 2018, pp. 12–21. ISBN: 978-1-4503-4925-3. DOI: 10.1145/3176349.3176377.

[43] Ran Yu et al. “Predicting User Knowledge Gain in Informational Search Sessions”. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. SIGIR ’18. New York, NY, USA: Association for Computing Machinery, June 2018, pp. 75–84. ISBN: 978-1-4503-5657-2. DOI: 10.1145/3209978.3210064.