Analyzing the Effects of Document's Opinion and Credibility on Search Behaviors and Belief Dynamics

Suppanut Pothirattanachaikul
Kyoto University
suppanut@db.soc.i.kyoto-u.ac.jp

Takehiro Yamamoto
University of Hyogo
t.yamamoto@sis.u-hyogo.ac.jp

Yusuke Yamamoto
Shizuoka University
yamamoto@inf.shizuoka.ac.jp

Masatoshi Yoshikawa
Kyoto University
yoshikawa@i.kyoto-u.ac.jp

ABSTRACT

To obtain accurate information through web searches, people have to search for information carefully. This study investigates how the search behaviors and decision outcomes of searchers were affected by the documents they encountered during their search process. We focus on two document factors: (1) *opinion* (consistent and inconsistent) with the searchers' beliefs prior to the search task, and (2) *credibility* (high and low). We conducted a user study in which 260 participants were asked to perform health-related search tasks while controlling a search result with different opinions and credibility levels. The results revealed that (i) the participants spent more effort searching by issuing more queries, when belief-inconsistent documents were presented; (ii) the documents' opinion and credibility affected their *belief dynamics*, (*i.e.,* how their beliefs changed after the search task); and (iii) their belief dynamics and search efforts had few relationships. These findings suggest that search engines could prevent users from polarization and thus, help them to obtain accurate information, by presenting documents that are inconsistent with users' beliefs on the higher-rank of the results.

KEYWORDS

Behavior Analysis, Credibility, Confirmation Bias

ACM Reference Format:
Suppanut Pothirattanachaikul, Takehiro Yamamoto, Yusuke Yamamoto, and Masatoshi Yoshikawa. 2019. Analyzing the Effects of Document's Opinion and Credibility on Search Behaviors and Belief Dynamics. In *The 28th ACM International Conference on Information and Knowledge Management (CIKM '19), November 3–7, 2019, Beijing, China.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3357384.3357886

1 INTRODUCTION

To obtain accurate information, previous studies have suggested that users should expose themselves to diverse information [12, 24, 25, 35] and carefully verify information by spending more time on a search task, checking the author information in a document during a search, *etc* [47, 49]. However, users are likely to exhibit confirmation bias, where they seek information to confirm their beliefs via web search engines [42–45]. For example, the users who believe that green tea extract is helpful for losing weight might select only positive information about the green tea extract, even though researchers argue that it is ineffective and harmful [4, 18]. Consequently, they are likely to refrain from diverse and careful searches for information.

Another line of research suggested that opinion and the quality of information affect people's confirmation bias and information behavior [7, 20, 38]. For example, Schwind et al. reported that people's confirmation bias could be mitigated by presenting information that was inconsistent with their preference [38]. Such researches imply that we would be able to design a search results ranking system that encourages people to search more carefully.

In this study, we focused on two *factors* of the document searchers encountered during their searches: (1) *opinion* that is either consistent or inconsistent with the searchers' beliefs prior to the search task, and (2) *credibility*, which represents the believability of the document. We examined the extent to which these factors encouraged or discouraged searchers from searching carefully. The aim of our study is to learn the characteristics of a ranking system that can encourage people to search more carefully by better understanding how these factors affect their search behaviors and *belief dynamics* (*i.e,* how their beliefs change after the search process).

More precisely, the purpose of the study is to answer the following research questions: **(RQ1):** How do opinion and credibility encourage or discourage searchers to search for information carefully?, **(RQ2):** How do opinion and credibility affect searchers' decision outcomes in terms of belief dynamics?, and **(RQ3):** Is there any relationship between search behaviors and belief dynamics?

The participants were required to use our search system to find the answer to four medical-related *yes-no* questions, such as *Are dairy products effective in improving hypertension?* Prior to the search task, they were asked to provide their beliefs in relation to the remedy, which was then defined as *prior belief*. We controlled the search results list by inserting a manually prepared document with a specific opinion and credibility level depending on the experiment condition and the participants' prior beliefs in the second rank of the list. The participants were also asked to provide their beliefs in relation to the remedy after the search task, which was then defined as *posterior belief*. By doing so, we were able to analyze the effects the opinion and credibility had on the search behaviors and belief dynamics.

CIKM '19, November 3–7, 2019, Beijing, China

ACM ISBN 978-1-4503-6976-3/19/11…$15.00
https://doi.org/10.1145/3357384.3357886

The main findings of the study are summarized as follows: (1) the participants spent more effort searching by issuing more queries when they encountered documents that were inconsistent with their prior beliefs; (2) they tended to alter their beliefs when encountering an opinion that was inconsistent with their prior beliefs while they tended to retain their beliefs when encountering an opinion that was consistent with their prior beliefs. Also, such an effect depended on the level of credibility; and (3) the behavioral differences we found in (1) could not be observed according to the participants' belief dynamics as only the types of opinion in the clicked documents were highly different. These results imply that search engines could prevent users from polarization and thus, help them obtain accurate information, by presenting documents that are inconsistent with users' beliefs on the higher-rank of the results.

2 RELATED WORK

2.1 Confirmation Bias

Confirmation bias is the phenomenon where people seek information that confirms their predispositions without considering contradictory opinions [17, 33]. The effect of confirmation bias can be explained by the defense mechanism, cognitive dissonance, in which people prefer supportive information over opposing information to avoid mental discomfort [6]. This kind of behavior may lead people to make disastrous decisions because they may seek the information that deviates from the truth, or they may overlook the risks of confirmatory information [15, 32, 42].

2.2 Credibility

In this study, we follow the definition of credibility by Fogg, who simply defined it as the *believability* of information and its sources [10]. Fogg and Tseng proposed a taxonomy on the credibility of web information. It is comprised of four categories based on items used to judge the credibility of information: presumed credibility, earned credibility, surface credibility, and reputed credibility [11]. We employed surface credibility to prepare documents with different credibility levels (See Section 3.3). To evaluate information carefully, credibility awareness is essential. Metzger et al. found that credibility awareness impacts information seeking behaviors in which people are likely to be a critical information consumer [28]. Therefore, we considered credibility to be a major factor that could encourage people to search for information more carefully.

2.3 Bias, Credibility, and Search Behaviors

Previous studies found that people's search behaviors and decisions can be affected either by the bias in the search results or by the credibility of the sources. As for the bias in the search results, White and his colleagues [42–45] found that people's confirmation bias can be exacerbated during searches; they found that people's decisions after searching were more biased towards a medical remedy was helpful even though there was evidence that it was not helpful. Moreover, they found that it was difficult to change people's beliefs after their decisions has been made. In addition to White's findings, Pogacar et al. [34] recently found that people were likely to make incorrect decisions when the search results were biased towards incorrect information. As for credibility, Kammerer et al. [19] found that people were more likely to spend less effort evaluating search results when the information sources seemed credible. Metzger [27] found that people were likely to believe visually aesthetic web information due to their limited time and cognitive capacity.

In order to mitigate confirmation bias and make more effective decisions, people should be exposed to alternative or opposing information that they might not have chosen previously [3, 14, 31, 41]. In line with this ideology, previous studies have found that recommending information inconsistent with the user's preference can stimulate divergent thinking and consideration of diverse information [24, 25, 37, 38].

Our study considered and extended these studies in the following ways. First, while the previous studies [24, 25, 37, 38] revealed that the people's confirmation bias can be mitigated by presenting information inconsistent with their preference, the detailed relationship between their search behaviors and decisions when encountered with inconsistent information during a search is still unclear. Second, we consider both the opinion of a search result and its credibility, which enables us to better understand their relationship. Third, we also consider the relationship between people's belief dynamics and their detailed search behaviors to better understand how their search behaviors relate to their decisions.

3 CONCEPT

In this section, we define a *critical thinker* who is expected to carefully search for information and consider different points of view. Next, we introduce *opinion* and *credibility* as the document factors investigated in this study. Finally, we describe how we prepared the documents based on these factors.

3.1 Critical Thinker in Web Search

Ennis defined critical thinking as the process of careful thinking to determine what to believe or react to. Ideally, critical thinkers should seek alternatives, consider other points of view, and be aware of their own beliefs [5]. Consistent with this definition, Yamamoto et al. [47–49], discussed the behaviors of critical thinkers in the context of web searching: They are expected to spend more time searching, issue more queries, browse more documents so as to make comparisons, and increase the evidence to support their decisions. More precisely, in the present study, we assumed that, compared to the non-critical thinkers, critical thinkers: (1) issue more queries, (2) click on more documents, (3) check deeper ranked result, (4) spend more time on browsing a document, (5) consider more documents as evidence that support their decisions, and (6) consider more reliable documents as evidence

The aim of our study is to understand the extent to which a document's opinion and credibility encourage or discourage the search behaviors described above (**RQ1**). We also investigate the effects document's opinion and credibility have on people's belief dynamics (**RQ2**) and its relationship with their behaviors (**RQ3**).

3.2 Factors Investigated in This Study

In this study, *opinion* and *credibility* are considered a document's major factors that affect users' search behaviors and belief dynamics. Opinion is the polarity of the document's content compared to the users' *prior beliefs*, beliefs held before performing a search. We investigated two types of opinions, *consistent*, where the polarity of

(a) Low Credibility Controlled Document

(b) High Credibility Controlled Document

Figure 1: Screenshots of our controlled documents with (a) low credibility and (b) high credibility (translated from Japanese). With the exception of visual styles, both controlled documents show the same content in their articles.

the content supports the users' prior beliefs and *inconsistent*, where the polarity of the content opposes the users' prior beliefs. Opinion is considered because it impacts users' information selection. While belief-inconsistent opinions encourage users to select diverse information, belief-consistent opinions exacerbate their confirmation bias [38].

Credibility is also taken into account in this study; it has been viewed as an important factor that influences the behaviors of system users [30]. In this study, we prepared *high* and *low* credibility documents separately. Previous studies proposed that people became less careful when information that seemed credible was presented and more careful when information that seemed less credible was presented [14, 19]. Therefore, we expected people to be less likely to search for information carefully when a high credibility document was presented and more likely to search for information carefully when a low credibility document was presented. Some studies have revealed that opinion and the quality of information are related [7, 20]. Thus, we also expected different combinations of opinions and levels of credibility to have different impacts on users' search behaviors and decision outcomes.

3.3 Controlled Documents Preparation

To investigate the effects different opinions and credibility levels have on peoples' search behaviors and belief dynamics, we manually prepared documents based on these factors. We refer to them as *controlled documents* hereinafter.

As for opinion, to prepare documents that were consistent/inconsistent with the users' prior beliefs, we manually prepared the content that supports/opposes to each search task. For example, for a search task that required a participant to find an answer for *Are dairy products effective in improving hypertension?*, we collected two types of articles from the web; ones that claimed that *dairy products are effective in improving hypertension*, and ones that claimed that *dairy products are NOT effective in improving hypertension*. The contents were selected from reliable sources that satisfied either of the following conditions: (1) the article explicitly refers to an academic paper or technical report presented at an academic conference, (2) the article is from a reliable medical knowledge base[1], and (3) the article in which author information is explicitly written. Table 1 shows the list of search tasks used in our study and the sources for preparing the controlled documents.

As for credibility, we developed the document layout in accordance with *surface credibility* [10]. Surface credibility is an evaluation criterion that the believability of information is judged based on appearance (as in, visual style). One typical example of surface credibility is that people are more likely to believe aesthetically appealing information [40]. We selected the surface credibility because it is easy for even people without an educational background to understand and apply [10, 29, 36] and many people apply surface credibility to evaluate web information [27]. As shown in Figure 1, the high credibility controlled document was designed to be clean, clear, and simple since the document with such components are likely be perceived as aesthetic [21, 22]. Meanwhile, the low credibility controlled document was intentionally designed as less aesthetically appealing than the high credibility ones.

To avoid situations in which the differences in their content would affect the participants' search behaviors and decision outcomes, both the high and low credibility controlled documents presented the same content. Also, as a document's URL plays an important role in judging surface credibility [10], we prepared the domain names for each credibility controlled document [2]. As we will cover in Section 5.1, the majority of the participants rated our high credibility documents as more credible than our low credibility ones. We prepared 16 controlled documents (4 search tasks × 2 opinion types × 2 credibility levels) for the user study described in the next section.

4 METHOD

We introduce the experimental design in this section and describe the experiment procedure in detail. We also introduce the search system that was tailored for this study and how we manipulated the search results. Finally, we outline how the participants were recruited and filtered from the analysis.

4.1 Experimental Design

The experiment was conducted with two independent within-subject variables: *opinion* (consistent/inconsistent) and *credibility* (high/low). It generated four experiment conditions which are C_1 (inconsistent, high credibility); C_2 (inconsistent, low credibility); C_3 (consistent, high credibility); and C_4 (consistent, low credibility). Our experiment was repeated in that each participant was assigned to all four conditions. For each condition, the participants were assigned to one of the four medical *yes-no* questions adopted from previous literature [49], because medical information tends to have reliable evidence for both supporting and opposing opinions. Also, the degree of belief can be measured by *yes-no* questions [42]. Table 1 details each question. The tasks were rotated according to the Graeco-Latin square design [23].

4.2 Procedure

First, participants were informed how their data would be used and asked for consent to use their data for research purposes. Only

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

[2]High Credibility controlled document: http://www.med-japan.com/ ,
Low Credibility controlled document: http://www.kenko-blog-life.org/ (The word "kenko" means healthy in Japanese)

Table 1: Search tasks used in this study. Sources of controlled documents are also shown.

Search Task		Sources	
ID	Questions	Supporting	Opposing
1	Are dairy products effective in improving hypertension?	http://bit.ly/2Tr2mYG (refers to an academic paper)	http://bit.ly/2TmjsHo (refers to an academic paper)
2	Are isoflavones effective in relieving high cholesterol?	https://nkbp.jp/2CNTjM5 (refers to a technical report)	http://bit.ly/2Tshjdc (medical knowledge base)
3	Is ginseng effective in improving dementia?	http://bit.ly/2COPciQ (refers to a technical report)	http://bit.ly/2CPDV21 (medical knowledge base)
4	Is homeopathy effective in improving asthma?	http://bit.ly/2TrKGfA (explicit author information)	http://bit.ly/2Tq7Z9B (medical knowledge base)

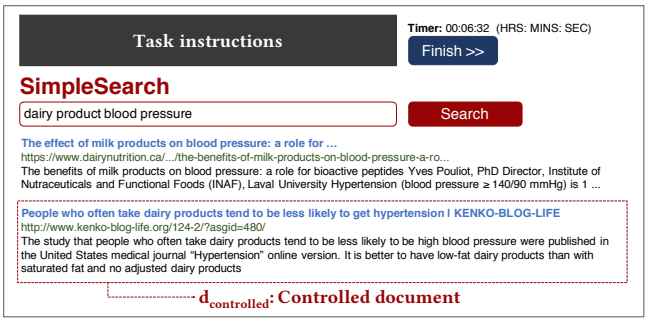

Figure 2: Screenshot of our search system. The controlled document was inserted at the second rank of search results.

participants who agreed to these conditions could participate in the training task. In the training task, they were asked to find the answer for the question *Is Chinese medicine effective for atopic dermatitis?* using our search system to get familiar with our search system. Note that search results were not manipulated during the training task.

After the training task, the participants were asked to complete four search tasks. Each search task was performed as follows. First, the participants were given the instructions that comprised the background of the symptom and a remedy for each search task. In order to avoid a situation in which the participants becoming biased because of the instructions, we tried to ensure the instructions were as neutral as possible. The following phrase is an example of the instructions (for task 1 in Table 1) that were presented to the participants:

Hypertension tends to be a disease of the heart and the vascular system. People with hypertension typically have a high mortality rate. Therefore, preventing and relieving hypertension is important. Suppose that you are currently having hypertension symptoms and the doctor tells you to lower your blood pressure. You have heard rumors that dairy products such as milk, yogurt, and cheese seem to be effective in improving hypertension. Therefore, in this task, please conduct a web search to examine whether dairy products are effective in improving hypertension.

After reading the instructions, the participants were asked to complete a pre-task questionnaire. First, the participants were asked to provide their beliefs in relation to the remedy by the question *Do you think dairy products are effective in improving hypertension?* with a four-point Likert scale (1: No; 2: Lean No; 3: Lean Yes; 4: Yes). This means that the participants had to decide their beliefs to be either of the two sides. A possible problem with using a four-point Likert scale will be discussed later in Section 6.2. We regarded the participants' scores for this question to be their *prior beliefs*. Second, the participants were asked to rate their prior knowledge about the symptom and remedy on a four-point Likert scale (1: Not at all; 2: A little; 3: Good; 4: Excellent).

Subsequently, the participants performed the search task by using our search system. The details of our search system are explained in Section 4.3. During the search task, they were asked to use our system and were prohibited from using other commercial search engines. Also, they were allowed to issue any queries to the system and click on any documents returned by the system. The participants were not limited by time constraints. Once the participants felt satisfied, they could click the finish button at the top right of the search page.

At the end of each search task, the participants were asked to complete a post-task questionnaire. They were first asked to provide their beliefs about the remedy again on the same four-point Likert scale. We regarded their scores for this question as their *posterior beliefs*. Next, the participants were asked to select the documents that contain evidence to support their posterior beliefs. We provided a list of the documents they clicked on during the search task. They were asked to select one or more documents among the list.

After the four search tasks were finished, we asked the participants who clicked on the controlled document(s) during the four search tasks to evaluate the credibility of the controlled document(s) with the question *How likely were you to believe this document?* (1: Not at all likely to believe it; 2: Likely not to believe it; 3: Likely to believe it; 4: Very likely to believe it) for each controlled document they clicked, as we wanted to ensure that the majority of the participants perceived the high credibility controlled documents as credible and the low credibility ones as less credible. Thereafter, they had to complete an exit questionnaire that required them to provide demographic information (gender, education level, and search engine familiarity). Note that the participants had the option to refrain from answering the demographic questions on gender and educational background. Finally, the participants were asked if anything had bothered them during the search task. This questionnaire was prepared so that we could ascertain who had noticed that we manipulated the search results by inserting a controlled document in the search results.

4.3 Search System

Figure 2 shows the screenshot of our search system. In order to help the participants get familiar with our search system easily, we imitated the user interface of a commercial search engine with the task instructions above the search box.

Figure 3 shows the overview of how the search results were generated. According to the query issued by the participant, the system first obtained 50 organic search results $d_1, d_2, \ldots, d_{50}$ via the Bing Web Search API[3]. The system then inserted the controlled document $d_{\text{controlled}}$ within the second rank of the search results and showed

[3]https://azure.microsoft.com/en-us/services/cognitive-services/

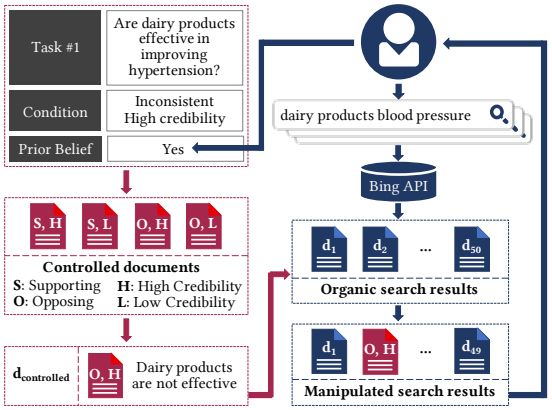

Figure 3: Search Result Manipulation Overview.

the top 50 search results, resulting in $d_1, d_{\text{controlled}}, d_2, \ldots, d_{49}$ being shown to the participant. Note that $d_{\text{controlled}}$ was selected based on the experiment condition and the participant's prior belief. For example, if the experiment condition was C_1 (inconsistent, high-credibility) and the participant's prior belief was either "Lean Yes" or "Yes", the system would present high-credibility $d_{\text{controlled}}$ containing information, which opposed the remedy (*i.e.,* there is *no* reliable evidence that the dairy products relieve the hypertension).

We expected that many of the participants would click on our controlled document during a search task as higher-ranked documents are likely to receive more attention [13, 16]. We decided not to insert the $d_{\text{controlled}}$ in the first rank of the search results because we want to avoid situations where the participants doubted that the search results were not generated by a usual search engine or they doubted that the search results were manipulated based on their prior beliefs. Each search result was comprised of a title, a snippet, and a URL. For $d_{\text{controlled}}$, we used its title and the first 100 words of its body text as a title and snippet, respectively. The document was displayed in a separate tab when it was clicked.

4.4 Participants

401 participants were recruited via Lancers.jp [4], a Japanese crowdsourcing platform, from July 10 to August 3, 2018. The median and standard deviation of the task completion time were 22 minutes and 26 minutes, 17 seconds, respectively. Since we used a crowdsourcing platform to conduct the experiment, controlling the quality is important. To do so, in the instruction page, we said that we might reject their submissions if we judged that they did not follow the task instruction, although we actually did not reject any submission of the 401 participants. Each participant received approximately $4 as compensation.

From the results of 401 participants, we removed 71 participants who clicked the controlled document less than two tasks. We further removed 24 participants who noticed the manipulation of the search results. Furthermore, we removed participants who used other search engines than ours, did not complete all the search tasks, or spent too much/too little time on the task. We collected the data of the search tasks in which a participant clicked on our controlled document and finally obtained 809 search tasks from 260 participants for analysis. The demographics of the 260 participants is shown in Table 2.

[4]https://www.lancers.jp/

Table 2: Demographics of participants.

Gender	n	Educational Background	n	Search Engine Familiarity	n
male	115	university educated	154	rarely use	1
female	141	not university educated	64	several times per week	22
N/A	4	N/A	42	once a day	45
				more than once a day	192

5 RESULTS

The results are reported as follows. First, we give the participants' perception of the credibility of our controlled document. Second, we examine the extent to which the controlled documents affected the participants' search behavior (RQ1). Third, we explore the extent to which the controlled documents affected the participants' belief dynamics (RQ2). Finally, we discuss the relationship between search behaviors and belief dynamics (RQ3). For the analyses regarding RQ1–RQ3, as we only used the data of the search task in which the participants clicked on our controlled documents, the data became unbalanced. The distribution of 809 search tasks in terms of the experimental conditions are C_1: 210, C_2: 191, C_3: 213, and C_4: 195. Therefore, we employed a linear mixed-effects regression analysis which is applicable for the unbalanced data [2, 8].

We constructed a linear mixed-effects regression model where *opinion*, *credibility* and their *interaction* were considered as fixed effects. In line with the recommendation of keeping the random effect structure maximal [1], the model initially included random intercepts and random slopes on participants and tasks. However, this model does not converge due to the data size. Therefore, we removed by-task random slopes because the tasks' variance is far smaller than participants' variance. Although removing the by-task random slopes decreases the generalizability of the model to the population of tasks, our model has high generalizability to the population of users [1, 46]. The final model we report contains random intercepts and slopes on participants and random intercepts on tasks. The formula for such a model [5] is $d \sim$ Opinion * Credibility + (1+ Opinion + Credibility | Participant) + (1 | Task), where d represents a dependent variable. We then performed the likelihood ratio test with the model and the *null model*, which only considers the random effects. The result is considered as significant if the model is statistically significant compared to the null model. The significance level in this study was set to 5%. Note that we applied log transformation to the temporal features, such as page dwell time.

5.1 Perception of Credibility

To ascertain whether the participants perceived our high (low) credibility documents as highly (low) credible, we calculated the average controlled documents' credibility score that had been obtained via the questionnaire for each credibility level separately. We found that high credibility controlled documents received higher average scores than low credibility controlled documents (high: $M = 2.89, SD = 0.80$; low: $M = 2.39, SD = 0.76$). The results

[5]The model is expressed in the syntax of lmer, a widely used mixed-effects fitting method contained in lme4 [2]

showed that the majority of the participants agreed that the high credibility controlled documents were credible and low credibility ones were less credible. To validate the significance of the results, we conducted the linear mixed-effects regression analysis by constructing a model where the credibility score was regarded as a dependent variable and *credibility* was regarded as a fixed effect. Note that the random effect structure for this model was the same as we explained above. We found the main effect of the *credibility* on the credibility score ($\beta = 0.27, SE = 0.03, t = 9.49, p < 0.001$). This result implies that different credibility levels affect the controlled documents' credibility score assessed by the participants.

5.2 Data Annotation

For the later analysis, we collected the documents (excluding controlled documents) clicked on by the participants during the search task (which consisted of 1,103 documents) and hired three graduate students majoring computer science as assessors to annotate them. First, we asked the assessors to annotate the types of opinions on the document as either *supporting*, *opposing*, *supporting and opposing*, or *irrelevant* to the search task. The Fleiss' kappa coefficient, which is used to measure the agreement among the assessors, for this annotation is 0.64, which indicates the substantial agreement among the assessors [9].

In addition to the types of opinions, we also asked the same assessors to annotate the reliability of the clicked documents. As we described in Section 3.1, we expected that people are likely to consider *reliable* documents as evidence when they are careful in their search process. To measure the reliability of the documents, we followed the four criteria of the JAMA benchmark [39]. The JAMA benchmark is the guideline that assesses the reliability of health care information on the web and has four criteria: (1) the document reveals its author (authorship), (2) updated date (currency), (3) references (attribution), (4) and sponsorship or ownership (disclosure). We asked the assessors to annotate whether the clicked document satisfies each of the four criteria. The Fleiss' kappa coefficients for the authorship, currency, attribution and disclosure were 0.72, 0.75, 0.58, and 0.68, respectively, all of which indicate the moderate to substantial agreement among the assessors. For later analyses, the annotation results were aggregated by majority vote, i.e., when at least two of the three assessors annotated the same label, it was considered as the gold standard label.

5.3 Search Behavior

As described in Section 3.1, our aim is to investigate to what extent a document's opinion and credibility encourage or discourage people to search for information carefully. The following search behaviors were analyzed for each participant during the search task:

- **# of Queries**: Number of queries issued in a search task.
- **# of Clicks**: Number of documents clicked on in a search task.
- **# of Consistent/Inconsistent Clicks**: Number of documents that were consistent/inconsistent with the participant's prior belief clicked on in a search task (excluding controlled documents).
- **Deepest Document Rank**: The lowest rank the participant clicked on.
- **Page Dwell Time**: Average time the participant spent on the documents.
- **SERP Dwell Time**: Average time the participant spent on the search results page.
- **Controlled Document Dwell Time**: Average time the participant spent on the controlled document.
- **Task Time Spent**: Amount of time the participant spent on the search task.

Note that the documents that contained both supporting and opposing opinions for the search task were not included in the analysis for both **# of Consistent Clicks** and **# of Inconsistent Clicks**.

We also analyzed the documents the participants selected as evidence that supported their posterior beliefs, which had been collected in the post-task questionnaire. We call these documents *evidence documents* hereinafter. The following metrics regarding the evidence documents were examined:

- **# of Evidence**: Number of evidence documents selected by the participant.
- **JAMA Benchmark**: Four numbers of evidence documents that satisfied (1) **Authorship**, (2) **Attribution**, (3) **Currency**, and (4) **Disclosure**, respectively.

5.4 Document's Opinion and Credibility on Search Behavior (RQ1)

To answer this research question, we performed a linear mixed-effects regression analysis with each participants' search behavior. Table 3 presents the results of the analysis. The "Fixed Effects" column in Table 3 shows the coefficient, standard error, t-statistic, p-value of each fixed effect in the model. This column also reports the model's goodness of fit when the model is statistically significant compared to the null model.

We found a main effect of the *opinion* on the number of queries ($\beta = 0.12, p < 0.05$). Table 4 shows the results of this behavior according to the types of opinion. The results in Table 4 revealed that the belief-inconsistent opinions encouraged the participants to issue more queries (inconsistent: $M = 2.70$, consistent: $M = 2.41$).

The results presented in Table 3 also revealed the main effect of *credibility* on the time spent on the controlled document ($\beta = 0.10$, $p < 0.01$). Table 5 shows the results of this behavior according to the types of credibility. The results in Table 5 revealed that the participants preferred spending time on high credibility controlled documents (high: $M = 19.67$, low: $M = 17.55$).

As for the interaction between opinion and credibility, we observed no significant interaction effect between the two. From this study, we did not observe the data that suggests the effect the different combinations of opinions and credibility levels had on the participants' search behaviors.

In essence, we found that the belief-inconsistent opinions encouraged the participants to issue more queries. A possible explanation for this result could be the participants who encountered the belief-inconsistent documents tried to issue more queries to find more information. However, we need a further study to understand this effect since its difference seems not large.

5.5 Belief Dynamics (RQ2)

To answer this research question, our analysis was based on the answer obtained from the belief-related questionnaires that were asked before and after the search task. As for belief dynamics,

Table 3: Mean (M) and standard deviation (SD) for each controlled document condition (C_1: inconsistent and high credibility, C_2: inconsistent and low credibility, C_3: consistent and high credibility, and C_4: consistent and low credibility) and statistical testing results of each search behavior of the model with different fixed effects (*: significance level at 0.001, **: 0.01, and *: 0.05). Coefficient (β), standard error (SE), t-statistic (t), and p-value (p) of each fixed effect are reported. The χ^2 and p-values of the model which is statistically significant compare to the null model are also reported.**

	Condition								Fixed Effects													
	C_1		C_2		C_3		C_4		Opinion				Credibility				Opinion × Credibility					
	M	SD	M	SD	M	SD	M	SD	β	SE	t	p	β	SE	t	p	β	SE	t	p	χ^2	p
# of Queries	2.67	2.00	2.74	2.19	2.32	1.65	2.51	1.90	0.12	0.05	2.57	*	-0.07	0.05	-1.39	0.17	0.02	0.04	0.49	0.62	9.12	*
# of Clicks	6.11	3.52	6.32	3.51	5.68	2.82	6.02	3.32	0.14	0.08	1.87	0.06	-0.07	0.08	-0.81	0.42	0.03	0.07	0.46	0.65		
# of Consistent Clicks	1.50	1.75	1.24	1.70	1.38	1.82	1.34	1.91	0.01	0.06	0.11	0.92	0.07	0.06	1.17	0.24	0.05	0.06	0.83	0.40		
# of Inconsistent Clicks	2.21	2.47	2.31	2.46	2.03	2.18	2.45	2.49	0.02	0.06	0.28	0.78	-0.10	0.08	-1.20	0.23	0.08	0.06	1.34	0.18		
Deepest Document Rank	11.25	9.05	12.40	11.00	10.99	8.69	12.08	10.59	0.15	0.25	0.61	0.54	-0.45	0.27	-1.69	0.09	0.07	0.23	0.29	0.77		
Page Dwell Time (sec)	25.40	21.91	26.25	23.92	28.00	31.40	28.80	32.70	-0.02	0.02	-1.28	0.20	0.00	0.02	0.21	0.83	0.02	0.02	1.45	0.15		
SERP Dwell Time (sec)	38.54	29.30	41.08	35.63	41.90	45.10	41.28	36.06	0.00	0.01	-0.04	0.97	-0.01	0.02	-0.73	0.47	-0.01	0.01	-0.88	0.38		
Controlled Document Dwell Time (sec)	17.10	18.00	18.10	35.50	22.20	52.30	17.00	21.40	-0.04	0.02	-1.61	0.11	0.10	0.03	3.30	**	0.01	0.02	0.42	0.67	11.80	**
Task Time Spent (sec)	413.79	319.97	419.88	270.30	414.62	354.36	401.68	252.58	0.02	0.01	1.34	0.18	-0.01	0.01	-0.54	0.59	0.00	0.01	-0.03	0.97		
# of Evidence	2.61	1.70	2.60	1.63	2.29	1.28	2.43	1.50	0.10	0.04	2.41	*	-0.02	0.04	-0.58	0.56	0.02	0.04	0.52	0.60		
Authorship	0.17	0.50	0.18	0.40	0.16	0.37	0.22	0.47	-0.01	0.02	-0.41	0.68	-0.01	0.01	-0.99	0.32	0.01	0.01	0.71	0.48		
Attribution	0.29	0.56	0.25	0.49	0.30	0.57	0.26	0.56	-0.01	0.02	-0.33	0.74	0.02	0.02	1.22	0.22	0.00	0.02	0.10	0.92		
Currency	0.72	1.11	0.79	0.89	0.63	0.76	0.76	0.92	0.03	0.03	0.95	0.34	-0.05	0.03	-1.68	0.09	0.02	0.03	0.60	0.55		
Disclosure	0.78	1.06	0.84	0.98	0.66	0.76	0.72	0.94	0.06	0.03	1.92	0.06	-0.03	0.03	-0.97	0.33	0.00	0.03	0.01	0.99		

Table 4: Participants' behavior according to opinion types.

	Inconsistent		Consistent	
	M	SD	M	SD
# of Queries	2.70	2.09	2.41	1.77

Table 5: Participants' behavior according to credibility level.

	High		Low	
	M	SD	M	SD
Controlled Document Dwell Time (sec)	19.67	39.26	17.55	29.20

in this study, we focused on the cases in which the participants changed (or did not change) their prior beliefs from one polar to the opposite polar after the search task. We refer to the cases in which the participants changed their prior beliefs from one polar to the opposite polar as they *altered* their beliefs, otherwise, *retained*.

We first performed a logistic mixed-effects regression analysis to investigate how opinion and credibility affected belief dynamics. Similar to Section 5.4, we constructed a logistic mixed-effects model where belief dynamics d was considered a binary dependent variable (1: alter, 0:retain). From the logistic mixed-effects regression analysis, we found the main effect of the document's *opinion* ($\beta = 0.71, SE = 0.10, z = 7.35, p < 0.001$) and *credibility* ($\beta = -0.21, SE = 0.09, z = -2.19, p < 0.05$) on the belief dynamics. This suggests that both opinion and credibility affected the participants' belief dynamics. Note that the model is statistically significant by the likelihood ratio test compared to the null model.

To further analyze the effects of opinion and credibility, Tables 6 (a)–(d) show the participants' beliefs after the search task (columns) compared to their beliefs prior to the search task (rows) with the different opinions and credibility conditions. Tables 6 (a) – (d), show that the participants were more likely to alter their beliefs from one polar to the opposite polar (①) when belief-inconsistent controlled documents were presented, especially from the *no* polar to *yes* polar, compared with the cases when belief-consistent controlled documents were presented (②). For example, when belief-inconsistent, high credibility controlled documents were presented, 71.43% (53.85% + 17.58%) of the participants who were "Lean No" prior to the search task altered their beliefs to either "Lean Yes" or "Yes" after the search task. On the contrary, as cells ③ and ④ show, the participants were more likely to retain their beliefs in the same polar when belief-consistent controlled documents were presented. These results imply that the participants who encountered a belief-inconsistent controlled document during a search process were more likely to alter their beliefs (especially when their beliefs were either "No" or "Lean No" prior to the search task). On the other hand, the participants who encountered a belief-consistent controlled document were more likely to retain their beliefs.

As for credibility, we found that the participants were less likely to retain their beliefs when belief-consistent, low credibility controlled documents were presented, (④), compared with cases when belief-consistent and high credibility controlled documents were presented (③). This result implies that the effect the belief-consistent controlled document had depended on its credibility level. Participants who encountered a belief-consistent controlled document with higher credibility were more likely to retain their beliefs.

Finally, as we can see from Tables 6 (a) and (b), we observed that the participants were likely to lean their beliefs towards the *yes*-polar after the search task when belief-inconsistent documents were presented. The possible explanation for this result is that the majority of the organic search results leaned towards *yes*-polar. Table 7 shows the distribution of the opinions in the documents clicked on by the participants, from which we can observe that most search results leaned towards the supportive information about the search task. This result implies that the distributions of the opinions in the search results would have an impact on the searchers' decision outcomes, as also reported by [34, 43].

An interesting finding was that such an effect could not be observed for the participants who encountered the belief-consistent opinions. Even the distribution of the documents' opinion leaned towards *yes*; as shown in Table 7, many participants who believed *no*

Table 6: Percentage of belief dynamics for each controlled document condition. (For each prior belief, the polar that the majority of participants leaned towards in posterior belief is highlighted in gray). The presentation of this table is inspired by the paper [42].

(a) C_1: Inconsistent, High Credibility (n = 210)

		Posterior Belief			
		No (n=21)	Lean-No (n=50)	Lean-Yes (n=104)	Yes (n=35)
Prior Belief	No (n =31)	16.13%	32.26%	25.81%	25.81%
	Lean No (n=91)	5.49%	23.08%	53.85%	17.58%
	Lean Yes (n=79)	13.92%	22.78%	50.63%	12.66%
	Yes (n=9)	-	11.11%	77.78%	11.11%

(b) C_2: Inconsistent, Low Credibility (n = 191)

		Posterior Belief			
		No (n=28)	Lean No (n=34)	Lean Yes (n=99)	Yes (n=30)
Prior Belief	No (n=31)	29.03%	19.35%	41.94%	9.68%
	Lean No (n=99)	10.10%	14.14%	56.57%	19.19%
	Lean Yes (n=56)	14.29%	25.00%	50.00%	10.71%
	Yes (n=5)	20.00%	-	40.00%	40.00%

(c) C_3: Consistent, High Credibility (n = 213)

		Posterior Belief			
		No (n=51)	Lean No (n=66)	Lean Yes (n=66)	Yes (n=30)
Prior Belief	No (n=47)	48.94%	31.91%	14.89%	4.26%
	Lean No (n=96)	25.00%	46.88%	27.08%	1.04%
	Lean Yes (n=65)	6.15%	9.23%	49.23%	35.38%
	Yes (n=5)	-	-	20.00%	80.00%

(d) C_4: Consistent, Low Credibility (n = 195)

		Posterior Belief			
		No (n=37)	Lean No (n=55)	Lean Yes (n=72)	Yes (n=31)
Prior Belief	No (n=37)	37.84%	37.84%	24.32%	-
	Lean No (n=99)	21.21%	34.34%	37.37%	7.07%
	Lean Yes (n=57)	3.51%	12.28%	45.61%	38.60%
	Yes (n=2)	-	-	-	100.00%

prior to the search task still retain their beliefs in *no* when the belief-consistent opinion was presented. This phenomenon might be caused by the effect of the participants' confirmation bias that was strengthened by belief-consistent opinions. Therefore, it was difficult to change their beliefs even when the majority of the search results contained the *yes*-polar information.

In summary, we found that both opinion and credibility affected the participants' belief dynamics. As for opinion, participants were likely to alter their beliefs from one polar to the opposite polar when a belief-inconsistent controlled document was presented. As for credibility, participants were likely to retain their beliefs when the belief-consistent, high credibility controlled document was presented.

5.6 Search Behaviors and Belief Dynamics (RQ3)

To explore the relationship between search behaviors and belief dynamics, we performed a logistic mixed-effects regression analysis. For each search behavior b_i investigated in Section 5.3, we constructed a logistic mixed-effects regression model M_i where the

Table 7: Distribution of the clicked documents' opinion for each task (excluding controlled documents) where TaskID refers to task questions in Table 1.

TaskID	Supporting	Opposing	Supporting & Opposing
1	87	9	5
2	50	5	1
3	52	1	2
4	33	20	3

Table 8: Mean (M) and standard deviation (SD) for each type of belief dynamics and statistical testing result of each search behavior with the belief dynamic as dependent variables (*: significance level at 0.05 ,*: significance level at 0.001). Coefficient (β), standard error (SE), and z-statistic (z) are also reported.**

	Belief Dynamics				Model				
	Alter		Retain						
	M	SD	M	SD	χ^2	p	β	SE	z
# of Consistent Clicks	0.70	1.25	1.83	1.97	81.94	***	-0.53	0.07	-7.53
# of Inconsistent Clicks	3.10	2.60	1.65	2.06	67.54	***	0.31	0.04	7.37
Authorship	0.14	0.37	0.21	0.48	5.36	*	-0.42	0.20	-2.03
Attribution	0.23	0.51	0.30	0.57	6.14	*	-0.36	0.16	-2.20

participant's belief dynamics d (0: retain, 1: alter) is considered as a dependent variable and b_i is considered as a fixed effect. We then evaluated the goodness of fit of M_i compared to the null model by using the likelihood ratio test and reported the results in Table 8 when M_i is statistically significant.

We found the main effect of the *number of consistent clicks* (β = -0.53, $p < 0.001$), *number of inconsistent clicks* (β = 0.31, $p < 0.001$), *authorship* (β = -0.42, $p < 0.05$), and *attribution* (β = -0.36, $p < 0.05$) on the belief dynamics. Table 8 shows the results of these behaviors according to the types of belief dynamics. Table 8 revealed that the participants who altered their beliefs were less likely to click on the documents that were consistent with their prior beliefs (alter: M = 0.70, retain: M = 1.83) and were more likely to click on the documents that were inconsistent with their prior beliefs (alter: M = 3.10, retain: M = 1.65).

For the evidence related behaviors, Table 8 revealed that the participants who retained their beliefs were more likely to submit the evidence that satisfies authorship (alter: M = 0.14, retain: M = 0.21) and attribution (alter: M = 0.23, retain: M = 0.30). Although we could not provide a clear explanation for this result, this could be explained by the psychological phenomenon called *motivated reasoning* [20] in which people spend their efforts collecting information that confirms their prior beliefs to maintain their beliefs in the face of information that challenges their beliefs. However, these behavioral differences seem smaller than those in the number of clicked documents that were consistent/inconsistent with the participants' beliefs.

As Table 8 shows, while we found that the participants issued more queries when the belief-inconsistent documents were presented (Section 5.4), such a behavioral difference was not found in terms of the belief dynamics. Only the types of opinion of the clicked documents were highly different. This implies that the participants' belief dynamics and their search effort had few relationships.

Table 9: Distribution of prior knowledge for each task where TaskID refers to task questions in Table 1.

TaskID	not at all	a little	good	excellent
1	168	28	6	0
2	154	42	18	0
3	190	7	1	0
4	178	12	3	2

6 DISCUSSION

In this section, the implication of the results and limitations of the study are discussed.

6.1 Implications

Our findings imply that search engines could mitigate the confirmation bias by learning the users' prior beliefs through their search histories and presenting documents that are inconsistent with their prior beliefs at the higher rank in the search results, as they are likely to be influenced by the higher-ranking documents [16]. By doing so, users can consider the different viewpoints and obtain accurate information. Recently, search engines have started summarizing the contradictory information about a query and presenting it as featured snippets to draw an attention from their users [6]. Presenting the belief-inconsistent information in the featured snippets might be helpful to encourage users to aware of their underlying confirmation bias.

As for credibility, we observed that the participants were likely to retain their beliefs when the highly credible and belief-consistent documents were presented (Section 5.5). This result implies that people might be more likely to believe the inaccurate content if it seems credible and confirms their prior beliefs. Recently, some major search engines encouraged content providers to follow design frameworks such as AMP, and search results that followed the framework would be quickly displayed to the users in a well-designed fashion[7, 8]. One concern is that such search results would likely to be considered credible by users, since documents with high usability and visually aesthetic are often considered credible [26, 50]. Therefore, both search engines and their users should be aware that such search results might strengthen users' confirmation bias.

6.2 Limitations

Our work has several limitations that should be acknowledged.

Effect of organic search results: It might be questioned whether the participants' behaviors and belief dynamics were affected by organic search results as well as by the controlled documents. Regarding the organic search results, we observed the results similar to White's [43], where people were likely to lean their beliefs towards the dominant polar in the search results, as shown in Table 7 that the majority of search results were leaning towards *yes*. However, we also observed the effect controlled documents had on their belief-dynamics; the participants who believed *no* prior to the task still retained *no* as their posterior beliefs when the controlled document that was consistent with their prior beliefs was presented. This means that the participants' behaviors and belief dynamics were affected by both the controlled documents and the organic search results. The current analyses reported in Section 5 cannot fully separate these two effects.

Noticing of the search result manipulation: We removed the participants who noticed our search result manipulation via the question in the exit questionnaire (Section 4.2). One limitation of such an approach is that we could not identify all the participants who noticed the manipulation. However, since we did not tell the participants the exact research questions investigated in the study, we still regarded the data of these participants as useful even if they felt strange about the presented search results.

Belief-related Scale: We used the 4-point Likert scale to measure the participants' prior and posterior beliefs, which means that we assumed they were able to decide either lean-yes or lean-no (Section 4.2). One problem with this assumption is that it would be difficult for participants to decide their beliefs (especially before starting the search task) when they had little knowledge about the topic; thus, their beliefs were weak. As shown in Table 9, most participants had little knowledge of the search tasks. However, we observed that the opinion affected the behaviors (Section 5.4) and belief dynamics (Section 5.5) of even such participants.

Generalizability: Our study cannot be generalized to other domains of search tasks. The question as to whether our findings are applicable to other domains remains open. For example, the political domain, where people are more likely to have strong prior beliefs and where complicated diverse opinions exist on the web. However, we think that the search tasks tackled in this study (*i.e.*, health-related search tasks that can be answered *yes* or *no*) are also an important potion of search tasks in our recent society.

7 CONCLUSION

In this study, we investigated the effects documents' opinion and credibility have on people's search behaviors and belief dynamics. The results were that the participants are likely to spend more effort in their search process by issuing more queries when the documents that were inconsistent with the participants' prior beliefs were presented. Furthermore, they tended to alter their beliefs when an opinion that was inconsistent with their prior beliefs was presented, and they tended to retain their beliefs when an opinion that was consistent with their prior beliefs was presented. Also, the effect depended on the level of credibility. Finally, the participants' belief dynamics and search efforts had few relationships; only the types of opinion in the clicked documents were different. In the future, we aim to analyze the search tasks in which the participants did not click on the controlled documents because it would be interesting to observe whether they were affected by the title or the snippet of the controlled document even though they did not click on it. We also aim to develop a new re-ranking algorithm to encourage searchers to perform careful information searches.

ACKNOWLEDGMENTS

This work was supported in part by JSPS KAKENHI Grant Numbers JP18H03494, 18KT0097, JP16H01756, 18H03243, and JP16H02906.

[6] https://www.blog.google/products/search/reintroduction-googles-featured-snippets/
[7] https://blog.google/products/search/search-results-are-officially-ampd/
[8] https://blogs.bing.com/Webmaster-Blog/September-2018/Introducing-Bing-AMP-viewer-and-Bing-AMP-cache/

REFERENCES

[1] Dale J. Barr, Roger Levy, Christoph Scheepers, and Harry J.Tily. 2011. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68, 3 (2011), 255–278.

[2] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-Effects models using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48.

[3] Cēline Buchs, Fabrizio Butera, Gabriel Mugny, and Cēline Darnon. 2004. Conflict elaboration and cognitive outcomes. *Theory Into Practice* 43, 1 (2004), 23–30.

[4] Naga P. Chalasani, Paul H. Hayashi, Herbert L. Bonkovsky, Victor J. Navarro, William M. Lee, and Robert J. Fontana. 2014. ACG clinical guideline: the diagnosis and management of idiosyncratic drug-induced liver injury. *The American Journal of Gastroenterology* 109 (2014), 950–966.

[5] Robert H. Ennis. 1987. *A taxonomy of critical thinking dispositions and abilities.* W H Freeman/Times Books/Henry Holt & Co. 9–26 pages.

[6] Leon Festinger. 1957. *A theory of cognitive dissonance.* Stanford University Press.

[7] Leon Festinger. 1964. *Conflict, decision, and dissonance.* Stanford University Press.

[8] Andy Field, Jeremy Miles, and Zoë Field. 2012. *Discovering statistics using R.* SAGE Learning.

[9] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378.

[10] Brian J. Fogg. 2003. *Persuasive technology using computers to change what we think and do.* Morgan Kaufmann Publishers.

[11] Brian J. Fogg and Hsiang Tseng. 1999. The elements of computer credibility. In *Proceedings of the 17th SIGCHI Conference on Human Factors in Computing Systems (CHI'99).* 80–87.

[12] Kelly R. Garrett and Paul Resnick. 2011. Resisting political fragmentation on the Internet. *Daedalus, Journal of the American Academy of Arts & Sciences* 140 (2011), 108–120.

[13] Zhiwei Guan and Edward Cutrell. 2007. An eye tracking study of the effect of target rank on web search. In *Proceedings of the 25th SIGCHI Conference on Human Factors in Computing Systems (CHI'07).* 417–420.

[14] Ivan Hernandez and Jesse Lee Preston. 2012. Disfluency disrupts the confirmation bias. *Journal of Experimental Social Psychology* 49 (2012), 178–182.

[15] Irving L. Janis. 1982. *Groupthink.* Cengage Learning.

[16] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. 2007. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transaction on Information Systems (TOIS)* 25, 2 (2007).

[17] Eva Jonas, Stefan Schulz-Hardt, Dieter Frey, and Norman Thelen. 2001. Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology* 80 (2001), 557–571.

[18] Tannis M. Jurgens, Anne M. Whelan, Lara Killian, Steve Doucette, Sara Kirk, and Elizabeth Foy. 2012. Green tea for weight loss and weight maintenance in overweight or obese adults. *Cochrane Database of Systematic Reviews* 12 (2012).

[19] Yvonne Kammerer, Ivar Bråten, Peter Gerjets, and Helge I. Strømsø. 2013. The role of Internet-specific epistemic beliefs in laypersons' source evaluations and decisions during web search on a medical issue. *Computers in Human Behavior* 29, 3 (2013), 1193–1203.

[20] Jonas T. Kaplan, Sarah I. Gimbel, and Sam Harris. 2016. Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Scientific reports* 6 (2016), 39589.

[21] Kristiina Karvonen. 2000. The beauty of simplicity. In *Proceedings on the 1st conference on Universal Usability (CUU'00).* 85–90.

[22] Kristiina Karvonen, Lucas Cardholm, and Stefan Karlsson. 2000. Cultures of trust: a cross-cultural study on the formation of trust in an electronic environment. In *Proceedings of the 3rd Nordic Workshop on Security (NordSec'00).* 89–100.

[23] Diane Kelly. 2009. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval* 3, 1–2 (2009), 1–224.

[24] Vera Q. Liao and Wai-Tat Fu. 2014. Can you hear me now?: mitigating the echo chamber effect by source position indicators. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW'14).* 184–196.

[25] Vera Q. Liao, Wai-Tat Fu, and Sri Shilpa Mamidi. 2015. It is all about perspective: an exploration of mitigating selective exposure with aspect indicators. In *Proceedings of the 33rd SIGCHI Conference on Human Factors in Computing Systems (CHI'15).* 1439–1448.

[26] Gitte Lindgaard, Cathy Dudek, Devjani Sen, Livia Sumegi, and Patrick Noonan. 2011. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 1 (2011), 1:1–1:30.

[27] Miriam J Metzger. 2007. Making sense of credibility on the web: models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology (JASIST)* 58, 13 (2007), 2078–2091.

[28] Miriam J. Metzger, Andrew J. Flanagin, Alex Markov, Rebekah Grossman, and Monica Bulger. 2015. Believing the unbelievable: understanding young people's information literacy beliefs and practices in the United States. *Journal of Children and Media* 9, 3 (2015), 325–348.

[29] Miriam J. Metzger, Andrew J. Flanagin, and Ryan B. Medders. 2010. Social and heuristic approaches to credibility evaluation online. *Journal of Communication* 60, 3 (2010), 413–439.

[30] Morten Moshagen, Jochen Musch, and Anja S. Göritz. 2009. A blessing, not a curse: experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics* 52 10 (2009), 1311–20.

[31] Charlan Nemeth, Kathleen Mosier, and Cyntia Chiles. 1992. When convergent thought improves performance: majority versus minority influence. *Personality and Social Psychology Bulletin* 18, 2 (1992), 139–144.

[32] Charlan Nemeth and John Rogers. 1996. Dissent and the search for information. *British Journal of Social Psychology* 35 (1996), 67–76.

[33] Raymond S. Nickerson. 1998. Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology* 2, 2 (1998), 175–220.

[34] Frances A. Pogacar, Amira Ghenai, Mark D. Smucker, and Chales L.A. Clarke. 2017. The positive and negative influence of search results on people's decisions about the efficacy of medical treatments. In *Proceedings of the 3rd ACM SIGIR International Conference on Theory of Information Retrieval (ICTIR'17).* 209–216.

[35] Paul Resnick, Kelly R. Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. 2013. Bursting your (filter) bubble: strategies for promoting diverse exposure. In *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW'13).* 95–100.

[36] David Robins, Jason Holmes, and Mary Stansbury. 2009. Consumer health information on the web: the relationship of visual design and perceptions of credibility. *Journal of the American Society for Information Science and Technology (JASIST)* 61, 1 (2009), 13–29.

[37] Christina Schwind and Jürgen Buder. 2012. Reducing confirmation bias and evaluation bias: when are preference-inconsistent recommendations effective - and when not? *Computers in Human Behavior* 28, 6 (2012), 2280 – 2290.

[38] Christina Schwind, Jürgen Buder, Ulrike Cress, and Friedrich W. Hesse. 2012. Preference-inconsistent recommendations: an effective approach for reducing confirmation bias and stimulating divergent thinking? *Computers & Education* 58 (2012), 787–796.

[39] William M. Silberg, George D. Lundberg, and Robert A. Musacchio. 1997. Assessing, controlling, and assuring the quality of medical information on the internet: caveant lector et viewor - let the reader and viewer beware. *Journal of the American Medical Association* 277, 15 (1997), 1244–1245.

[40] Elizabeth Sillence, Pam Briggs, Peter Harris, and Lesley Fishwick. 2006. A framework for understanding trust factors in web-based health advice. *International Journal of Human-Computer Studies* 64, 8 (2006), 697–713.

[41] Cass R. Sunstein. 2017. *#Republic divided democracy in the age of social media.* Princeton Press.

[42] Ryen W. White. 2013. Beliefs and biases in web search. In *Proceedings of the 36th SIGIR Conference on Research and Development in Information Retrieval (SIGIR'13).* 3–12.

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

[44] Ryen W. White and Ahmed Hassan. 2014. Content bias in online health search. *ACM Transactions on the Web (TWEB)* 8, 4 (2014), 25:1–25:33.

[45] Ryen W. White and Eric Horvitz. 2015. Belief dynamics and biases in web search. *ACM Transactions on Information Systems (TOIS)* 33, 4 (2015), 18:1–18:46.

[46] Bodo Winter. 2013. Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint arXiv:1308.5499* (2013).

[47] Takehiro Yamamoto, Yusuke Yamamoto, and Sumio Fujita. 2018. 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).* 963–972.

[48] Yusuke Yamamoto and Takehiro Yamamoto. 2018. Query priming for promoting critical thinking in web search. In *Proceedings of the 3rd ACM SIGIR Conference on Human Information Interaction & Retrieval (CHIIR'18).* 12–21.

[49] Yusuke Yamamoto, Takehiro Yamamoto, Hiroaki Ohshima, and Kawakami Hiroshi. 2018. Web access literacy scale to evaluate how critically users can browse and search for web information. In *Proceedings of the 10th ACM Conference on Web Science (WebSci'18).* 97–106.

[50] Ping Zhang and Na Li. 2004. Love at first sight or sustained effect? The role of perceived affective quality on users' cognitive reactions to information technology. In *Proceedings of the 25th International Conference on Information Systems (ICIS'04).* 283–296.