Investigating the Influence of Featured Snippets on User Attitudes

Bink, Markus; Schwarz, Sebastian; Draws, Tim; Elsweiler, David

Citation (APA)
Bink, M., Schwarz, S., Draws, T., & Elsweiler, D. (2023). Investigating the Influence of Featured Snippets on User Attitudes. In *CHIIR 2023 - Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (pp. 211-220). (CHIIR 2023 - Proceedings of the 2023 Conference on Human Information Interaction and Retrieval). ACM. https://doi.org/10.1145/3576840.3578323

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Investigating the Influence of Featured Snippets on User Attitudes

Markus Bink
University of Regensburg
Regensburg, Germany
markus.bink@student.ur.de

Sebastian Schwarz
University of Regensburg
Regensburg, Germany
sebastian.schwarz@student.ur.de

Tim Draws
Delft University of Technology
Delft, The Netherlands
t.a.draws@tudelft.nl

David Elsweiler
University of Regensburg
Regensburg, Germany
david.elsweiler@ur.de

ABSTRACT

Featured snippets that attempt to satisfy users' information needs directly on top of the first search engine results page (SERP) have been shown to strongly impact users' post-search attitudes and beliefs. In the context of debated but scientifically answerable topics, recent research has demonstrated that users tend to trust featured snippets to such an extent that they may reverse their original beliefs based on what such a snippet suggests; even when erroneous information is featured. This paper examines the effect of featured snippets in more nuanced and complicated search scenarios concerning debated topics that have no ground truth and where diverse arguments in favor and against can legitimately be made. We report on a preregistered, online user study ($N = 182$) investigating how the *stances* and *logics of evaluation* (i.e., underlying reasons behind stances) expressed in featured snippets influence post-task attitudes and explanations of users without strong pre-search attitudes. We found that such users tend to not only change their attitudes on debated topics (e.g., school uniforms) following whatever stance a featured snippet expresses but also incorporate the featured snippet's logic of evaluation into their argumentation. Our findings imply that the content displayed in featured snippets may have large-scale undesired consequences for individuals, businesses, and society, and urgently call for researchers and practitioners to examine this issue further.

CCS CONCEPTS

• **Information systems → Search interfaces**.

ACM Reference Format:
Markus Bink, Sebastian Schwarz, Tim Draws, and David Elsweiler. 2023. Investigating the Influence of Featured Snippets on User Attitudes. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '23), March 19–23, 2023, Austin, TX, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3576840.3578323

1 INTRODUCTION

Search engine result page (SERP) characteristics such as the ranking and presentation of search results can influence users' interactions with those results [2, 21, 33, 42] and the outcomes of search sessions [1, 8, 16]. For example, viewpoint-biased search result rankings have been shown to lead to the *search engine manipulation effect* (SEME), a phenomenon whereby users without strong pre-existing viewpoints change their attitudes following whichever viewpoints most high-ranking search results express [1, 15, 16, 36]. More recent work has demonstrated that even just the presence of a *featured snippet* – a first-ranked and highlighted search result that aims to answer users' information needs directly – can induce such biased attitude change. Direct answers in featured snippets tend to draw more attention than regular results, reduce search time, and increase user satisfaction [48]. Users performing health-related searches perceive the information presented in featured snippets differently from other results, with information in featured snippets being judged as more credible, irrespective of its accuracy [6]. Moreover, featured snippets containing incorrect information can lead to users holding a false belief post-task, even when the outcome contradicts their initial viewpoint [6].

Despite these findings, two important aspects surrounding biased attitude change from featured snippets are currently unclear. First, the influence of featured snippets on user attitudes has been demonstrated in the health domain [6] but the extent of this effect, i.e., whether it also occurs for other debated topics (e.g., school uniforms), is not known. Second, because earlier work has used simple viewpoint taxonomies (e.g., looking only at stances or perceived credibility), it is still unclear to what degree user attitudes are affected in this context (i.e., whether users only follow the stance expressed by the featured snippet or also adopt its arguments, often referred to as *logics of evaluation* or simply *logics* [3, 11]). Uncovering in which situations and how drastically users may change their attitudes following the content they see in featured snippets can inform the development of a safer and more trustworthy web search experience for users. That is why, in this paper, we investigate whether biased featured snippets can affect user attitudes on commonly debated topics on multiple viewpoint dimensions. Our work is guided by two research questions:

RQ1. Do stances expressed in featured snippets affect the attitudes of web search users without strong pre-existing viewpoints?

RQ2. Do logics of evaluation expressed in featured snippets affect the attitudes of web search users without strong pre-existing viewpoints?

To address these research questions, we conducted a preregistered user study[1] that presented participants with SERPs related to topics for which they did not hold strong stances (Section 3). Participants could interact with the SERPs, each of which included a controlled featured snippet, as they would see when using a regular search engine (i.e., including visiting the actual web pages). Studying pre- and post-task user attitudes regarding both stances and logics, our results show that participants tended to not only change their attitudes on debated topics (e.g., school uniforms) following whatever stance a featured snippet expresses but also incorporate the featured snippet's logic of evaluation into their argumentation (Section 4). We discuss the limitations and implications of our work in Section 5. All supplementary material (e.g., task screenshots, data, and analysis code) is openly available at https://osf.io/fneu6.

2 RELATED WORK

Our research builds on existing knowledge concerning user biases in web search and how the presentation of search results (incl. featured snippets) can affect user behaviour; as well as work in the communication sciences that has recently impacted how viewpoints are represented and measured in human information interaction.

2.1 Biased Search Results and Search Behaviour

Web search commonly serves as a platform for opinion-forming related to debated topics [8, 19, 28] but a multitude of data-related, algorithmic, and (cognitive) user biases can influence users' search behavior and the outcomes of search sessions [2, 45]. Search results of popular search engines may often be biased in favor of particular viewpoints [13, 37, 42], types of people [32], or subtopics [18, 35]. Most likely due to a combination of data and ranking biases induced by the query or ranking algorithm at hand [18, 43], SERPs generally also contain more positive (i.e., query-affirming) results [42, 43] and often promote results that are already popular [10, 34].

Biased SERPs can lead to undesired consequences when cognitive user biases such as the *position bias* (i.e., users' tendency to pay more attention to high-ranking results [22]) and *exposure effects* (i.e., users' tendency to adopt the majority viewpoint among the results they consume [15]) come into play. A prominent example of an undesired effect on users in this context is the *search engine manipulation effect* (SEME), where users – especially those without strong pre-existing viewpoints – change their attitudes on a topic (e.g., an upcoming election or the efficacy of medical treatment) following whichever viewpoints are most prominent on the SERPs they see [1, 15, 16, 36] without necessarily being aware of it [20]. For factual web search, where users seek answers on scientifically answerable topics, Bink et al. [6] recently demonstrated that users even tend to change their attitudes in accordance with single featured snippets on top of the SERP – irrespective of the viewpoint distribution among the remaining search results. Our work is similar to that of Bink et al. [6] but looks at debated topics that do not have any clear scientific basis and where arguments in favor and against can legitimately be made.

While the above biases represent the focal point of our work, we note that users may additionally exhibit other irrational behavior. For example, users may be affected by the *confirmation bias*, i.e., seeking information that confirms their pre-existing beliefs and disregard contradicting information [24, 30, 38], or *anchoring* [41], i.e., judging results in comparison to those they have already seen [4, 40]. Furthermore, a user's pre-task topical knowledge [2, 5, 23, 44] and search expertise [46] can also impact on search outcomes.

2.2 Presentation of Search Results

The way search engine results are presented can affect users' perceptions of search results and the outcome of searches. This is reflected in theoretical contributions such as *information scent theory* [7] and has been validated empirically using click-through data [21, 33]. When result snippets are short, do not feature query terms, or include complex URLs, the probability of the associated document being viewed decreases [10]. The way search snippets are perceived and interpreted can vary significantly from user to user, with users basing their judgments on different components of listings [23]. It has been shown, however, that providing additional information, e.g., about a site's popularity or other meta-data, can affect how it is perceived by users [39, 50]. Moreover, recent research has highlighted the particularly powerful impact that featured snippets can have. Studies show that featured snippets attract more attention than regular result snippets and increase users' satisfaction [48]. Featured snippets also decrease engagement with the SERP in general [9, 47, 48], often leading the user to take the viewpoint expressed in the featured snippet [6].

2.3 Viewpoint Representation

Research in interactive information retrieval typically represents viewpoints on search topics in a binary fashion (e.g., [6, 23, 36]), that is, as either against or in favor of a given topic (e.g., school uniforms). Although such simple stance labels may be easy to obtain via crowdsourcing [29] or even automatic methods [12, 26], recent research has argued that they are an overly simplistic representation of viewpoints on debated topics [11]. Recent work has already begun to use more nuanced ordinal [14, 15] or continuous scales [27] to label stances expressed in documents. These approaches, however, still do not provide information about the underlying *reasons* behind viewpoints. Draws et al. [11], inspired by research from communication science [3], proposed a two-dimensional framework for viewpoint representation that complements stance with *logics of evaluation* to allow for more in-depth viewpoint analyses. Logics of evaluation, or simply *logics*, comprise seven categories that each broadly reflect a potential reason behind a stance: *inspired* (referring to what is true or divine), *popular* (referring to what is popular), *moral* (referring to what is social or fair), *civic* (referring to what is legal or accepted), *economic* (referring to what is profitable or creates value), *functional* (referring to what works), and *ecological* (referring to what is sustainable and natural). For example, someone with a stance *in favor* of school uniforms may argue using a *functional* logic by saying that school uniforms lead to better grades among students. Any argument, independent of topic or stance, can be classified into one of these seven logic categories.

In this work, we extend previous research in two key ways: (1) we examine the impact of featured snippets in a previously unstudied task scenario where users search to derive an opinion and (2) employ the viewpoint representation framework proposed by Draws et al. [11] to achieve a more nuanced understanding of how featured snippets influence users' viewpoints.

3 EXPERIMENTAL SETUP

To investigate the influence of featured snippets on users' attitude change regarding multiple viewpoint dimensions (i.e., stance and logics of evaluation), we performed a preregistered, web-based, randomized controlled trial. The study employed a between-subjects design with 12 conditions (stance x logic - see Table 1). Each condition represented a different SERP with search results related to one of three different debated topics and with a featured snippet as the top result that expressed one of two different stances and one of seven different logics.

3.1 Hypotheses

The stance of featured snippets has been shown to affect users in medical look-up tasks [6] (see Section 2). We predicted that this effect would transfer to search task scenarios where users have no strong preconceived attitudes and diverse arguments in favor and against can legitimately be made.

Hypothesis 1 (H1): The stance expressed in a featured snippet influences users' post-search attitude towards the debated topic the snippet refers to.

Featured snippets can change people's minds on health topics [6] and receive high levels of user interaction and attention [49]. We thus predicted an effect on the participants' attitudes based on the logic expressed in the featured snippet.

Hypothesis 2 (H2): The logic expressed in a featured snippet influences users' post-search attitude towards the debated topic the snippet refers to.

3.2 Materials

Our study makes use of several resources and performing the study necessitated design decisions for the SERPs and featured snippets.

3.2.1 Data. For our study, we used two openly available datasets containing search results (including their URL, title, and snippet) on eight different debated topics. The first data set (available at https://osf.io/yghr2) contains search results retrieved from two popular search engines for the topics *school uniforms*, *intellectual property rights*, and *atheism* [13]. All documents in this dataset are expert-annotated with viewpoint labels in the two-dimensional format proposed by Draws et al. [11]; including stance and logics (i.e., inspired, popular, moral, civic, economic, functional, ecological). The second dataset (available at https://osf.io/6tbvw), collected via the *Bing* API, contains search results for the topics *obesity as a disease*, *zoos*, *cell phone radiation safety*, *bottled water*, and *social networking sites*. While stance annotations for this dataset were obtained by the original authors via crowdsourcing, we manually added logics of evaluation for selected search results in this data set,
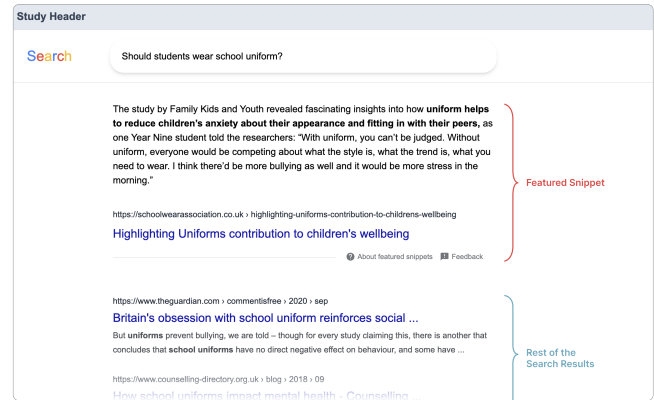
Figure 1: SERP with featured snippet and snippet for the query `Should students wear school uniform?`.

following the protocol from Draws et al. [11]. From these datasets, we selected topics (from the eight available ones) such that:

(1) viewpoints were balanced (i.e., mean stance for the documents for that topic was close to zero);
(2) documents were available with sufficient credibility (i.e., to remove a potential confounding variable);
(3) a spread of logics were available for our study.

We thus used search results from three topics for our user study:

- **Obesity** – Claim: *Obesity is a disease.*
- **Intellectual Property Rights** – Claim: *Intellectual property rights should exist.*
- **School Uniforms** – Claim: *Students should have to wear school uniforms.*

3.2.2 Featured Snippets. To create featured snippets for our user study, we manually extracted paragraphs from our data that could be used as such snippets. One of the authors then assigned a logic label to each paragraph that contained only a single logic. We finally selected 12 from these single-logic paragraphs for our user study such that each logic was represented at least once and the 12 paragraphs referred to a variety of topics. To ensure that logics could be applied consistently, the final 12 paragraphs were relabelled after a period of two weeks. 11 of the 12 featured snippets were labelled identically (Cohen's Kappa = .90). A language model (specifically *deepset/roberta-base-squad2* available through the Hugging Face community[2]) identified an answer within the paragraph that was highlighted in bold similar to a typical featured snippet on e.g., *Google* (see Figure 1). If the model was unable to find a realistic answer that could be used to bold the text, as is the case with realistic featured snippets, a suitable section was manually selected.

3.2.3 SERPs. For each of the three selected topics, we created four SERPs (i.e., resulting in a total of 12 different SERPs). Each SERP consisted of one featured snippet (ranked first) and nine additional search results, displayed in a ranked list similar to SERPs from current search engines. Per topic, we had selected four search results (i.e., two *in favor* and two *against*, with diverse logics) that would be used as featured snippets (see Section 3.2.2). Table 1 shows

---

[2]https://huggingface.co/deepset/roberta-base-squad2

an overview of the 12 different featured snippets, which eventually represented the 12 between-subjects conditions of our experiment.

To control confounding effects, we filled the remainders of each of the 12 SERPs with equal numbers of *in favor* and *against* results that were all rated as credible.[3] The results were ordered in alternating stance, i.e., if the featured snippet was *in favor*, the next result would be *against*, followed by *in favor*, and so on. All four search results per topic that were used as featured snippets always took positions 1 to 4 on the SERPs, with only the highest-ranked one of them being displayed as a featured snippet and the others as regular search results. Which result would be displayed as featured snippet varied with conditions (see Table 1). The six remaining results per SERP did not have strong stances (i.e., on a seven-point Likert scale ranging from *extremely against* to *extremely in favor*, they had been labelled as *somewhat against* or *somewhat in favor*). These six results continued in the alternating order of *in favor* and *against* stances. We did not control the logics that were featured in the search results in positions 2-10 as this was not practically possible but aimed for a good mix of logics on every SERP.

3.3 Variables

3.3.1 Independent Variables.

- **Stance**: Binary variable indicating the stance expressed in the featured snippet (*against* or *in favor*).
- **Logic**: Nominal variable indicating the logic of evaluation expressed in the featured snippet [3, 11]:
  - *Inspired*: speaking to what is true, divine, and amazing
  - *Popular*: speaking to what is popular or what people want
  - *Moral*: speaking to what is social, fair, and moral
  - *Civic*: speaking to what is legal, accepted, and conventional
  - *Economic*: speaking to what is profitable and creates value
  - *Functional*: speaking to what works
  - *Ecological*: speaking to what is sustainable and natural

3.3.2 Dependent Variables.

- **Attitude change**: difference between the measured pre- and post-attitude (i.e., users' stance; each measured on a seven-point Likert scale randing from strongly disagree; -3; to strongy agree; 3). Pre-search attitudes were all -1, 0, or 1 as we only tested users without strong pre-existing attitude for a topic (see following subsection). This meant that attitude change could range between [-4;4].
- **Logic Adoption**: binary variable indicating whether a user has adopted the logic expressed in the featured snippet for themselves post-search.

3.3.3 Descriptive and Exploratory Variables.
We use these variables to conduct descriptive and exploratory (but no confirmatory) analyses.

- **Time**: how much time participants spent on their search
- **Education**: participants' level of education

---

[3]We manually evaluated the credibility of search results using guidelines for credibility assessment developed in another project that is, as of this writing, still under review.

Figure 2: Screen that was shown to participants after they completed their search on the SERP. Participants were promted to update their stance on the respective topic, as well as could see their initial pre-task justification. If any new argumentations came to mind that were not previously mentioned, participant's could use the text field below.

3.4 Procedure

After reading about the study aims, process, and implications of their participation, participants signed a consent form. They then followed four steps to complete their participation:

(1) Participants provided basic demographic information. In addition, they indicated their stance on the three topics in our study. If any of their stances was one of the center three options on a seven-point Likert scale for at least one of the topics, participants were directed to step two. Otherwise, their participation ended with full compensation. We decided to **only include participants without strong pre-existing viewpoints** as they are particularly vulnerable to undesired effects such as SEME [15, 16].

(2) Subjects were randomly assigned to one of the topics for which they had a mild or neutral stance (i.e., a pre-search attitude in the center three options on the Likert scale). They were then asked to justify their position and reinforce it with any arguments in an open text field. These arguments are referred to as *justifications* in the remainder of the paper. We kept this voluntary because we did not want to push participants to give an explanation, which could lead to forced justifications.

(3) Within their topic, participants were randomly assigned to one of the four available featured snippets and saw the corresponding SERP (see Table 1). They were free to interact with the results as they would with a typical web search engine. Search result links to the different websites were clickable to allow for a natural interaction with the SERP.

(4) Participants gave their (potentially) updated attitude on the topic after finishing the search and were asked to enter any novel arguments (justifications) they now subscribe to in another open text field (see Figure 2).

Table 1: Featured snippets used in each of the 12 user study conditions. There are four featured snippets per topic (i.e., two in favor and two against) that express different logics in their direct answers. Each participant in our user study only saw one of these featured snippets on its corresponding SERP.

Topic	Stance	Logic	Featured Snippet Excerpt
Intellectual Property Rights	In favor	economic	Intellectual property (IP) contributes enormously…
Intellectual Property Rights	In favor	civic	Intellectual property is important to our daily lives…
Intellectual Property Rights	Against	moral	This is why around 100 countries, led by India and…
Intellectual Property Rights	Against	economic	Open and non-exclusive licensing unleashes the…
School Uniforms	In favor	popular	The study by Family Kids and Youth revealed…
School Uniforms	In favor	inspired	There's a number of reasons why most schools…
School Uniforms	Against	moral	Uniform policies reinforce gender and racial…
School Uniforms	Against	inspired	Picking out my own clothes and being free to…
Obesity	In favor	functional	Calling obesity what it is, "a disease," can help direct…
Obesity	In favor	ecological	Which brings me to the second point: Once…
Obesity	Against	civic	"We cannot say just because you are obese you …
Obesity	Against	inspired	Doctors should be required to tell patients a blunt…

3.5 Participants

A pre-study power analysis for an independent-samples t-test (see following subsection) using the software $G*Power$[17] (i.e., specifying the default medium effect size of $d = 0.5$, a significance threshold of $\alpha = 0.025$, and a desired power of 0.8) revealed that 156 participants would be required. We recruited a total of 256 participants from the *Prolific* platform (https://www.prolific.co) with each participant being paid £0.60 (i.e. £9.00 per hour). Since only people with mild pre-existing viewpoints were eligible to partake, only 182 people were able to complete the full study.[4]

3.6 Post-Processing and Statistical Analysis

We tested **H1** by conducting an independent samples t-test with the featured snippets' stances (e.g., either in favor or against) as independent and users' attitude change as dependent variable. To analyse the influence on users' change in logic (**H2**), we first annotated the participants' justifications using the logic framework proposed in [11] (see Section 4.3). This analysis only looked at those participants who did not already include their assigned featured snippet's logic of evaluation prior to searching (i.e., 99 out of the 182 participants who completed the study), as it was impossible for the other participants to consider adding the featured snippet's logic. We performed a binomial test on the logic adoption of those 99 participants, testing the null hypothesis that the chance of a user adopting the logic expressed in a featured snippet is $\frac{1}{7}$ (i.e., reflecting a uniform distribution considering that there are seven different logics that can apply to any topic). Because we tested two hypotheses in this study, we applied a Bonferroni correction to our significance threshold, thus reducing it to $\frac{0.05}{2} = 0.025$.

---

[4]While 182 is larger than the 156 suggested by the power analysis, it was difficult to control the number of eligible participants due to the filtering process. The effect sizes of our analyses indicate that the extra power is unproblematic (see Section 4).

4 RESULTS

Participants' ages ranged from 18 to 66 years ($M = 28.91, SD = 9.10$) with 69 identifying as female and 113 as male. The majority of participants (65.38%) were highly educated with 80 holding a bachelor's degree, 38 a master's degree, and one a doctoral degree. Sixty-two participants mentioned having at least a high school diploma and the remaining participants had a lower level of education.

4.1 Descriptive Statistics and Initial Exploration

We present initial analyses to understand how participants behaved. The most commonly allocated topic was *Obesity*, which was shown to 45.4% of participants. 29.3% of participants received *School Uniforms* and 25.2% were presented a SERP associated with *Intellectual Property Rights*. This uneven distribution is due to the fact that participants were randomly assigned to one of the topics on which they held a mild viewpoint. If this was the case for only one topic, the participant was assigned to that topic accordingly. Participants were presented with a near even split of against (47.2%) and in favor (52.8%) stances in featured snippets. Of the 182 participants who completed the study, 61 (33.5%) held a completely neutral view (i.e. a stance of 0 measured on a seven-point Likert scale), 35 (19.2%) a slightly against and 86 (47.2%) a slightly in favor stance towards their assigned topic. On average, participants provided pre-search justifications containing 27.43 words ($SD = 23.38$), with post-search justifications being only 15.96 words long ($SD = 16.27$). Pre- and post justification lengths were fairly similar across conditions.

Overall, the median time to complete tasks was 3.72 minutes ($IQR = 3.43 minutes$). This varied slightly across topics with the *Obesity* topic having the longest sessions ($Mdn = 3.94, IQR = 3.34$) followed by *School Uniforms* ($Mdn = 3.90, IQR = 4.74$) and *Intellectual Property Rights* ($Mdn = 3.35, IQR = 3.08$). In line with previous studies investigating direct answers or featured snippets [9, 48], few participants (17.58%) clicked on results. If results were clicked, however, there was a strong trend towards the featured snippet, which received more than half of all click-throughs (51.02%, see Figure 3). Results at position 7 and 8 received no clicks at all.
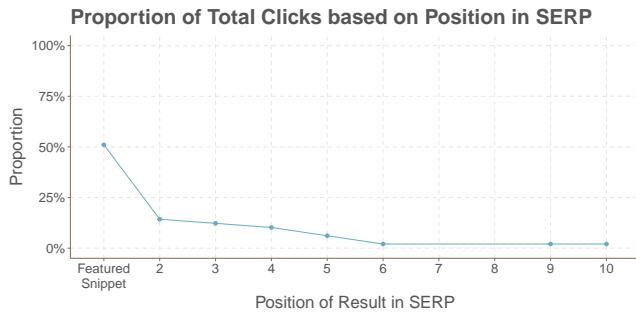
Figure 3: Relative click proportions over the ranks. Users behaved similarly across conditions, with a positional bias with regards to the featured snippet.

The justifications participants provided pre- and post-search were annotated for logics of evaluation using the framework proposed by Draws et al. [11] (see Section 4.3). On average, participants' justifications included 1.31 different logics pre-task ($SD = 0.88$) compared to 0.76 different logics post-task ($SD = 0.86$). There were fewer post-task logics since we asked people to add any additional arguments that they did not mention in the pre-task. The fact that the mean number of logics applied is less than one illustrates that not every SERP resulted in new justifications being won.

Figure 4a depicts the distribution of pre-task logics present in participants' justifications for featured snippets that contained a moral justification with an against stance and Figure 4b depicts the equivalent after the participant completed their search. Figures 4c and 4d likewise show the stances both pre- and post-task, respectively. These plots provide initial evidence that interacting with the SERPs impacted on the participants' argumentation. The moral logic is particulary more prominent post-task than is the case pre-task. Similarly, while pre-task stances had a clear positive lean, this is no longer the case post-task. This is a further hint that the featured snippets were influential.

4.2 RQ1: Influence of Featured Snippet Stance on Web Search Users' Attitude

To answer **H1** (i.e., whether the stance expressed in a featured snippet influences users' post-search attitudes toward a debated topic), we first computed participants' attitude change by subtracting their pre-task stance towards the respective topic from the post-task stance. For example, a participant with a pre-task stance of 1 and a post-task stance of 3 (measured on a seven-point Likert scale) experienced an attitude change of 2. **H1** is true if people's attitude change is in line with the stance expressed in the featured snippet. A t-test indicated a significant difference between attitude changes for featured snippets with *in favor* stances ($M = 0.55, SD = 1.09$) and featured snippets with *against* stances ($M = -0.28, SD = 1.30$; $t(180) = -4.68, p < .001, d = .70$), suggesting – in line with **H1** – that participants indeed changed their attitudes in accordance with the stances expressed in featured snippets.

4.3 RQ2: Influence of Featured Snippet Logic on Web Search Users' Justifications

To answer **H2** (i.e., whether the logic expressed in the featured snippet influences users' post-search justifications), each participant was first coded in a binary fashion as having adopted the logic expressed in the featured snippet or not. One author did this by blindly (i.e., without seeing other participant data) examining each participant's pre- and post-task viewpoint justification texts and annotating them for logics of evaluation; and finally coding participants depending on whether the logic of the featured snippet they saw appeared in their post-search justification. Entries were excluded from this labelling process if:

(1) they did not provide any pre- or post-task justification
(2) an explanation was given with no clear logic (i.e. if people gave an explanation like *"I don't know"*)
(3) their justification matched that provided by the featured snippet, but they had provided the same explanation pre-task (i.e. before seeing the featured snippet)
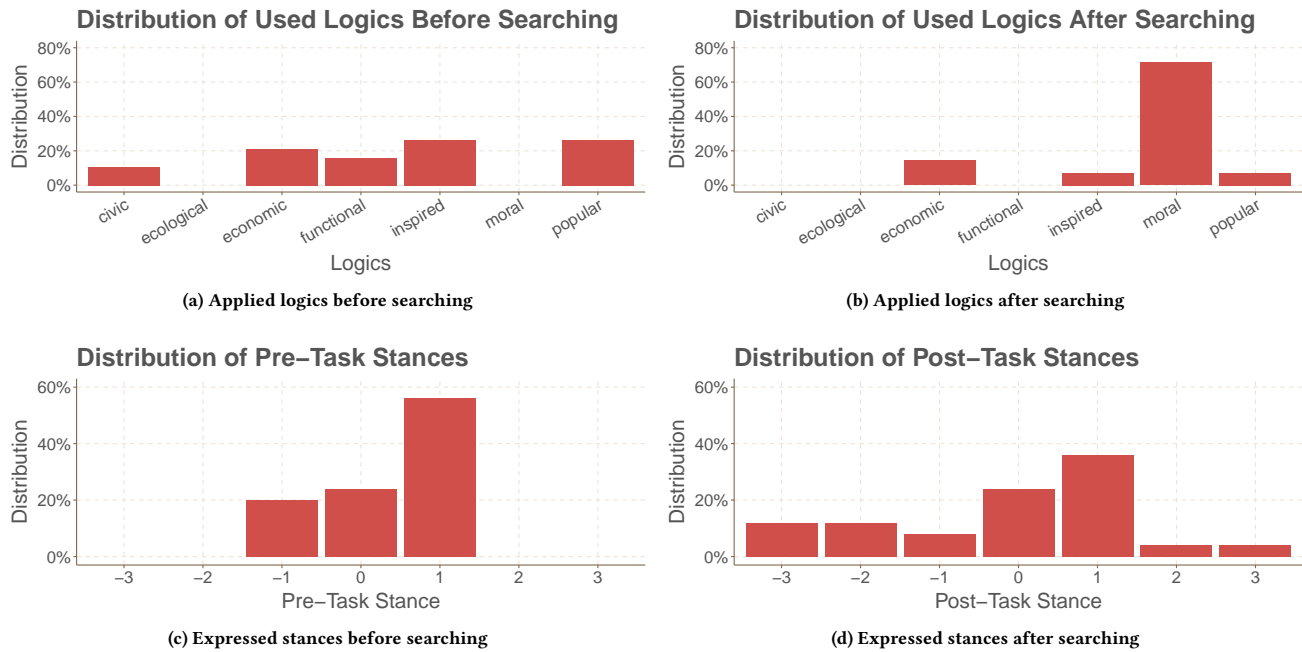
To ensure reliability, the data all each post-task explanations were re-coded after a duration of two weeks. Reliability was calculated using Krippendorff's alpha for multi-label nominal-scales since each justification could be labelled with one or more of the seven logics available. Krippendorff's α was .784 which is sufficient to draw conclusions [25, p.356].

In 32 of the 99 cases that remained after this filtering process (32.3%), participants reported a post-search justification that included the logic expressed in the featured snippet they had seen. To test whether this adoption of logics from the featured snippets was extraordinary, we assumed that there is a one in seven or 14.2% chance of any argument provided matching the featured snippet's logic (i.e., as there are seven logics). A binomial test indicated that, across conditions, participants were significantly more likely than chance to use the featured snippet's logic in their post-search justification (32.3%) than the expected 14.2% ($p < .001$).

4.4 Inspecting Justifications Qualitatively

The analyses in the previous two sub-sections were quantitative in nature. Here, we examine participants' stance justifications in more detail by presenting clear examples. Our aim in this qualitative analysis is to contextualise the results presented thus far.

During the logic-annotation process, it became clear that participants were not simply copying or parroting the arguments made in featured snippets. There were several indications, indeed in most of the responses, that the participants had invested considerable thought and effort into the process. There were post-task arguments provided where it was obvious that people had interacted with results. We saw evidence of learning, e.g., "*[the]search results revealed that wearing uniforms actually induce sexual harassment for female[s], which is really troubling*" **(U126)**, and the introduction to new, previously unconsidered, perspectives, e.g. "*Medical IPs and patents in general are something I did not consider. I believe that medical IPs should be forced to be leased to other producers (for a percentage of the profits of course) to combat artificial scarcity and monopolies (e.g. Insulin prices in the US).*" **(U079)**. Moreover, new information seems to have caused at least some participants to reflect on their original stance, e.g., "*After reading the google search,*

(a) Applied logics before searching

(b) Applied logics after searching

(c) Expressed stances before searching

(d) Expressed stances after searching

Figure 4: Distribution of logics and stances before and after participants were exposed to a SERP showing a featured snippet containing a moral logic that is against the presented topics (Intellectual Property Rights and School Uniforms, see Table 1).

I do not think I agree …" **(U081)**, and "*The text that was presented to me slightly changed my opinion on the subject. When I initially wrote my answer, I immediately thought of artistic creation. But what they presented to me was related to intellectual property on products that are indispensable to society, and I think that governments must intervene to guarantee a common good.*" **(U074)**.

Some participants made reference to specific documents that had influenced their thinking. For example, "*I found the Globe and Mail article interesting where it discussed how our bodies respond differently to overeating/binge eating …*" **(U146)**. Other participant entries revealed that participants had interacted with the presented results in a critical fashion. Some explanations refer to credibility judgements made about results, including the featured snippet, and suggest that these judgements had played into their arguments: "*The snippet at the top of the search results did not seem like it came from a reputable source.*" **(U168)**, "*I actually found that many reliable sources seem to think obesity is in fact not a disease. I did not gain any new reliable information that told me otherwise.*" **(U168)**. In this second example, it seems that the participant was weighing up the quantity of what they perceived to be reliable sources and this eventually led to their post-search stance.

There were also cases where participants demonstrated reflection but the information provided was insufficient to change their minds: "*I think I have expressed my opinion, search results don't influence my thoughts about copyright*" **(U040)**, "*Now I see more benefits for this. But we need more long term studies to know for sure*" **(U040)**, "*Not at all from what I just read I see that this is a two sided argument, there are some people who believe it is a disease and those who refuse to believe that it is*"**(U040)**. One participant expressed that search

results alone would not be enough for them to construct a view or change their perspective: "*… those searches, are just search topics. They won't change my mind, without further investigation*" **(U046)**, "*I will keep my opinion even after everything that I read*"**(U046)**.

Thus, it seems that participants were neither robotically stealing arguments from search results, nor were they superficially completing tasks just to receive remuneration with as little effort as possible. Our results suggest that participants took the task seriously and expended effort into considering the information presented, sometimes even being critical of the quality and sources. Yet, despite this effort, the findings presented in Sections 4.2 and 4.3, demonstrate that the SERPs shown influenced not only their post-task stance, but also the justifications they made for their stance.

4.5 Exploratory Analysis

To better understand our results, we conducted exploratory analyses concerning time, education level, and change in viewpoints.

4.5.1 Time. Given that featured snippets have been found to reduce interaction with the SERP and search time [48], we assume that participants with shorter search times are more likely to adopt the featured snippets' line of argumentation both in logic and stance. Participants were split into two groups by a median split based on their time on task (i.e., participants who searched for a time-period shorter than the median and those who searched for longer than the median). The adoption rate of the featured snippets logic did not differ significantly between participants who spent more time searching (38.0 %) and those who spend less time (26.5 %), ($\chi^2(1) = 1.01, p = .315$). Additionally, participants who spent more

time searching did not significantly change their stance to a greater degree (based on the absolute difference of post- and pre-task stance, $M = 0.82, SD = 0.90$) compared to participants who spend less time ($M = 0.86, SD = 1.01; t(180) = -0.23, p = .816$).

4.5.2 Educational Background. Looking at participants' educational backgrounds, highly educated participants (i.e. Bachelors, Masters or Doctorate) adopted the featured snippets logic slightly more often (33.3 %) than people with other educational backgrounds (30.6 %). Again, however, the difference was not significant ($\chi^2(1) = 0.003, p = .951$). A similar picture is also present with the number of logics expressed in their pre-task explanations. More highly educated participants did not express significantly more logics ($M = 1.35, SD = 0.85$) compared to participants with other educational backgrounds ($M = 1.25, SD = 0.94, t(97) = -0.53, p = .591$). The same was also true with respect to whether or not participants followed the featured snippets stance. This was true for 25.2 % of those participants who reported a higher level of education compared to 22.2 % for participants with other educational backgrounds.

4.5.3 Viewpoints. Among the participants who had mild pre-task stances (either -1 or 1, against or in favor respectively; rather than neutral pre-task stances), 53.7 % changed their viewpoints after searching. Those participants who were presented with conflicting viewpoints in the featured snippet (e.g., they were in favor of the topic in question but the featured snippet was against), changed their viewpoint in accordance with the featured snippet's stance in 14 % of cases. Around a third (32.2 %) of participants from the same groups (i.e., pre-task attitudes of -1 or 1) *strengthened* their initial viewpoint regardless of the presented information on the SERP. If participants followed the featured snippet's stance, *in favor* post-task stances were more extreme ($M = 1.86, SD = 1.49$) than *against* ones ($M = 0.35, SD = 2.32$). Of the participants who initially held a neutral viewpoint, 55.7 % changed their post-task viewpoint (in either direction) and 29.5% of these changed their viewpoints in accordance with the featured snippets' stance.

We finally looked at the influence of the number of logics people used in their pre-task explanations by splitting participants into two groups by using a median split based on the number of logics participants used in their pre-task justifications. Participants who applied more logics in their explanations did not significantly change their stance as much ($M = 0.94, SD = 1.03$) compared to participants who did not express many logics ($M = 1.25, SD = 0.77, t(97) = 1.14, p = .256$).

5 DISCUSSION

The presented study investigated the influence of featured snippets with a biased stance and controlled logic of argumentation on the post-task attitudes of searchers. To answer **RQ1**, while the reported post-task stances were by no means polemic, there is a clear pattern of movement to the stance of the featured snippet. The results show a significant effect that post-task, participants changed their attitudes in accordance with the stances expressed in featured snippets (Section 4.2). With respect to **RQ2**, the participants were also more likely than would be expected by chance alone to adopt the logic of argumentation used in the featured snippet in their own post-search justifications (Section 4.3). The exploratory analyses

we performed post-experiment did not reveal any demographic or behavioral trends. In other words, the findings seem to hold regardless of educational background or the time spent interacting with the search results.

Qualitatively inspecting the justifications participants provided for their stances revealed the subtle nature of the influence involved. The participants evidenced considerable thought and effort and often explained why they were convinced by what they had seen on the SERPs and indeed why not. Sometimes this involved being introduced to a perspective they had not considered (as in the examples above for **(U074)** and **(U079)** or arguments being made by credible sources (e.g. in the examples from **(U146)** and **(U168)**. Nevertheless, despite effort and critical evaluation on the part of the participants, there is clear evidence of movement towards the direction and argumentation of the featured snippet, regardless of what the stance was and how it was argued.

5.1 Implications

Our results build on past work to form a troubling picture: a highly motivated and, to a large extent, well-educated sample of participants adopt the stance and argumentation from a highlighted featured snippet after searching. This has implications for both search engines and the research community.

The picture painted by these and past findings underlines the power of search engines and begs the question of how they should wield this power. The results ask of modern search engines to take responsibility and make conscious decisions about how they want to handle viewpoint diversity etc.; thus moving away from the "our goal is just to provide relevant results" paradigm. Currently, some search engines, such as Google or Bing, provide users with a single biased answer to debated topics. Figure 5 depicts featured snippets by Google for each of the topics used in the study. These examples show a clear bias in one direction even for topics where a legitimate argument can be made in either direction. The last example by *britannica.com* especially stands out since even though the article presents arguments both for and against school uniforms, the corresponding featured snippet only gives a one-sided view on the topic. Given that featured snippets reduce engagement with SERPs [9, 47, 48] and increase satisfaction [48], it seems that users stick to the first viewpoint encountered and may not consider others, even if valid arguments could be made for both sides.

Thus, our findings imply that the content displayed in featured snippets may have large-scale undesired consequences for individuals, businesses, and society, and urgently call for researchers and practitioners to examine this issue further. Below we highlight topics for future work that would help understand the issue more clearly as well as help find solutions.

5.2 Limitations

Before highlighting future work, it is important to acknowledge that our research is subject to some limitations. For one, participants were only able to view one of three fixed topics and not individually chosen ones. Although many more debated topics exist and some participants might be more interested in other ones, this allowed us to control the effect the topic might have on users' search outcome by creating SERPs with specific characteristics such
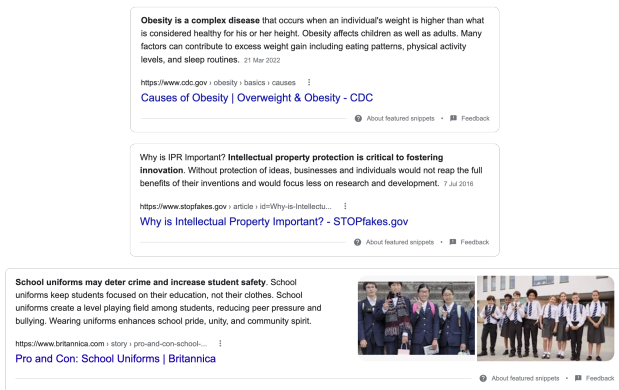
Figure 5: Examples of real-world featured snippets that appear for the same topics used in the study. All explanations in the featured snippet are biased toward a specific direction - either against or in favor of the topic in question.

as the featured snippets' logic and stance. Additionally, our preselection made sure that participants had no strong opinions for at least one of the topics which left participants open to explore new viewpoints as would be the case in a more naturalistic scenario. While it may have been better to filter participants such that they have no preconceived opinion on all three topics and then assign a topic randomly, this would have been infeasible since hardly any of the recruited participants would have been eligible. Even with unlimited resources, we believe filtering so many participants may have introduced unintended sampling bias.

Given that participants did not hold strong stances on the topics they were assigned, it is possible that they had not given the topic much thought or were not particularly knowledgeable on the topic. This may have made them easier to sway, indeed past work has suggested that in such cases users may be particularly vulnerable to undesired effects such as SEME [15, 16]. This effect may have been compounded by the fact that we only presented credible documents.

Finally, our analysis of participants' logic adoption (see Section 4.3) is limited in two ways. First, we could only consider the data of about half of participants (as all other participants had already stated the featured snippet's logic pre-search). This was necessary as we wished to specifically look at whether users adopt particular logics but did not account for some cases, e.g., users *dropping* a logic from their justification post-search. Second, we had to make a strong assumption in all logics being equally likely candidates for post-search. We note here, however, that any null hypothesis (e.g., considering the pre-search instead of a uniform logic distribution) would reflect a strong and potentially unfounded assumption here. We believe that the uniform logic distribution was the most unbiased of all these options and that our analyses, in combination with descriptive statistics, present a convincing picture.

6 CONCLUSION AND FUTURE WORK

The presented study investigated how the stances and logics of evaluation expressed in featured snippets influence post-task attitudes

and justifications of users without strong pre-search attitudes. Featured snippets were manipulated such that they showed an answer with either an *in favor* or *against* stance and included one of seven logics in the explanation. Our findings show that users were not only significantly affected by the answers' stances but also adapted the featured snippets' logic into their own post-search argumentation. This evidences that the influence of featured snippets is not only present in the health-domain (as previous research showed [6]) but also extends to other, more general debated topics such as school uniforms. We also show, that participants did not just blindly follow the answer presented in the featured snippet but put significant effort and consideration into their justifications. Given these results, it begs the question of whether featured snippets should take side on debated topics or whether there are situations in which it is reasonable to present the user with a singular viewpoint.

Future work should explore if and how to present users with answers without strong biases. One way could be to present users with more than a singular answer. While still making use of the positive qualities of featured snippets (e.g. showing an answer that can be found in the full document), this approach would leave the user the opportunity to compare possible answers and decide accordingly. Some search engines (e.g., *Bing*) already do this for some questions but it is not yet clear if this has an effect on users' opinion forming and, if so, how users are influenced. Another approach might be to inform users about the biases present in the SERP or to design search interfaces that help educate users to identify biases themselves. Such boosting approaches are appealing since they benefit the user beyond the current context of the intervention [31].

There are many other open questions that research still needs to address. For example, the SERPs studied in our work were manufactured and presented in controlled environments and it would be fascinating to study these situations naturalistically (e.g., as in Ortloff et al. [31]). Moreover, we controlled our experiment to study the effects on people without strong pre-existing attitudes. An intuitive next step would be to see the effects, if any, on users who already have a strong stance. In summary, while search engines are a vital informational and learning tool, the subtle biasing effect their results can have on our opinions and behavior is worrying and requires further study.

REFERENCES

[1] Ahmed Allam, Peter Johannes Schulz, Kent Nakamoto, et al. 2014. The impact of search engine selection and sorting criteria on vaccination beliefs and attitudes: two experiments manipulating Google output. *Journal of medical internet research* 16, 4 (2014), e2642.

[2] Leif Azzopardi. 2021. 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* (Canberra ACT, Australia) *(CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 27–37. https://doi.org/10.1145/3406522.3446023

[3] Christian Baden and Nina Springer. 2014. Com(ple)menting the news on the financial crisis: The contribution of news users' commentary to the diversity of viewpoints in the public debate. In *European Journal of Communication 29*. European Journal of Communication, 529–548. https://doi.org/10.1177/0267323114538724

[4] Marcia J Bates. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online review* (1989).

[5] Suresh K Bhavnani. 2002. Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In *CHI'02 extended abstracts on human factors in computing systems*. 610–611.

[6] Markus Bink, Steven Zimmerman, and David Elsweiler. 2022. Featured Snippets and Their Influence on Users' Credibility Judgements. In *ACM SIGIR Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR*

'22). Association for Computing Machinery, New York, NY, USA, 113–122. https://doi.org/10.1145/3498366.3505766

[7] Stuart K. Card, Peter Pirolli, Mija Van Der Wege, Julie B. Morrison, Robert W. Reeder, Pamela K. Schraedley, and Jenea Boshart. 2001. Information Scent as a Driver of Web Behavior Graphs: Results of a Protocol Analysis Method for Web Usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. 498–505.

[8] Noel Carroll. 2014. In search we trust: exploring how search engines are shaping society. *International Journal of Knowledge Society Research (IJKSR)* 5, 1 (2014), 12–27.

[9] Lydia B. Chilton and Jaime Teevan. 2011. Addressing People's Information Needs Directly in a Web Search Result Page. In *Proceedings of the 20th International Conference on World Wide Web (WWW '11)*. 27–36.

[10] Charles L. A. Clarke, Eugene Agichtein, Susan Dumais, and Ryen W. White. 2007. The Influence of Caption Features on Clickthrough Patterns in Web Search. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Amsterdam, The Netherlands) (*SIGIR '07*). Association for Computing Machinery, New York, NY, USA, 135–142. https://doi.org/10.1145/1277741.1277767

[11] Tim Draws, Oana Inel, Nava Tintarev, Christian Baden, and Benjamin Timmermans. 2022. Comprehensive Viewpoint Representations for a Deeper Understanding of User Interactions With Debated Topics. In *CHIIR '22: ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22)*. ACM SIGIR Conference on Human Information Interaction and Retrieval, 135–145.

[12] Tim Draws, Karthikeyan Natesan Ramamurthy, Ioana Baldini Soares, Amit Dhurandhar, Inkit Padhi, Benjamin Timmermans, and Nava Tintarev. 2023. Explainable Cross-Topic Stance Detection for Search Results. In *CHIIR '23: ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '23)*. ACM SIGIR Conference on Human Information Interaction and Retrieval.

[13] Tim Draws, Nirmal Roy, Oana Inel, Alisa Rieger, Rishav Hada, Mehmet Orcun Yalcin, Benjamin Timmermans, and Nava Tintarev. 2023. Viewpoint Diversity in Search Results. In *European Conference on Information Retrieval*. Springer.

[14] Tim Draws, Nava Tintarev, and Ujwal Gadiraju. 2021. Assessing Viewpoint Diversity in Search Results Using Ranking Fairness Metrics. In *ACM SIGKDD Explorations Newsletter (ACM SIGKDD '21)*. Association for Computing Machinery, New York, NY, USA, 50–58.

[15] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. 2021. This Is Not What We Ordered: Exploring Why Biased Search Result Rankings Affect User Attitudes on Debated Topics. In *SIGIR '21: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 295–305.

[16] Robert Epstein and Ronald E. Robertson. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. In *Proceedings of the National Academy of Sciences 112*. PNAS, E4512–E4521. https://doi.org/doi.org/10.1073/pnas.1419828112

[17] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.

[18] Ruoyuan Gao and Chirag Shah. 2020. Toward creating a fairer ranking in search engine results. *Information Processing & Management* 57, 1 (2020), 102138.

[19] Lisa Gevelber. 2018. *It's all about 'me'—how people are taking search personally*. Technical Report. Technical Report. https://www.thinkwithgoogle.com/marketing-strategies ....

[20] Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. 2020. A Think-Aloud Study to Understand Factors Affecting Online Health Search. In *CHIIR '20: Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. CHIIR, 273–282. https://doi.org/doi.org/10.1145/3343413.3377961

[21] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2017. Accurately interpreting clickthrough data as implicit feedback. In *Acm Sigir Forum*, Vol. 51. Acm New York, NY, USA, 4–11.

[22] 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 Trans. Inf. Syst.* 25, 2 (apr 2007), 7–es. https://doi.org/10.1145/1229179.1229181

[23] Markus Kattenbeck and David Elsweiler. 2019. Understanding credibility judgements for web search snippets. *Aslib Journal of Information Management* (2019).

[24] Joshua Klayman and Young-Won Ha. 1987. Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review* 94, 2 (1987), 211.

[25] Klaus Krippendorff. 2018. *Content analysis: An introduction to its methodology*. Sage publications.

[26] Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. *ACM Computing Surveys (CSUR)* 53, 1 (2020), 1–37.

[27] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P Gummadi, and Karrie Karahalios. 2019. Search bias quantification: investigating political bias in social media and web search. *Information Retrieval Journal* 22, 1 (2019), 188–227.

[28] Dana Mckay, Stephann Makri, Marisela Gutierrez-Lopez, Andrew MacFarlane, Sondess Missaoui, Colin Porlezza, and Glenda Cooper. 2020. We Are the Change That We Seek: Information Interactions During a Change of Viewpoint. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (Vancouver BC, Canada) (*CHIIR '20*). Association for Computing Machinery, New York, NY, USA, 173–182. https://doi.org/10.1145/3343413.3377975

[29] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*. 31–41.

[30] Raymond S. Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology* 2, 2 (1998), 175–220.

[31] Anna-Marie Ortloff, Steven Zimmerman, David Elsweiler, and Niels Henze. 2021. The Effect of Nudges and Boosts on Browsing Privacy in a Naturalistic Environment. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 63–73.

[32] Jahna Otterbacher, Jo Bates, and Paul Clough. 2017. Competent men and warm women: Gender stereotypes and backlash in image search results. In *Proceedings of the 2017 chi conference on human factors in computing systems*. 6620–6631.

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

[34] Sandeep Pandey, Kedar Dhamdhere, and Christopher Olston. 2004. WIC: A General-Purpose Algorithm for Monitoring Web Information Sources. In *Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30* (Toronto, Canada) (*VLDB '04*). VLDB Endowment, 360–371.

[35] Sachin Pathiyan Cherumanal, Damiano Spina, Falk Scholer, and W Bruce Croft. 2021. Evaluating Fairness in Argument Retrieval. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 3363–3367.

[36] Frances A Pogacar, Amira Ghenai, Mark D Smucker, and Charles LA Clarke. 2017. The positive and negative influence of search results on people's decisions about the efficacy of medical treatments. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. 209–216.

[37] Cornelius Puschmann. 2019. Beyond the bubble: Assessing the diversity of political search results. *Digital Journalism* 7, 6 (2019), 824–843.

[38] Alisa Rieger, Tim Draws, Mariët Theune, and Nava Tintarev. 2021. This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media*. 189–199.

[39] Julia Schwarz and Meredith Morris. 2011. Augmenting Web Pages and Search Results to Support Credibility Assessment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. 1245–1254.

[40] Milad Shokouhi, Ryen White, and Emine Yilmaz. 2015. Anchoring and Adjustment in Relevance Estimation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15)*. 963–966.

[41] Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185, 4157 (1974), 1124–1131.

[42] Ryen White. 2013. Beliefs and biases in web search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. 3–12.

[43] Ryen W. White. 2014. Belief Dynamics in Web Search. *J. Assoc. Inf. Sci. Technol.* (2014), 2165–2178.

[44] Ryen W White, Susan T Dumais, and Jaime Teevan. 2009. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the second ACM international conference on web search and data mining*. 132–141.

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

[46] Ryen W White and Dan Morris. 2007. Investigating the querying and browsing behavior of advanced search engine users. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. 255–262.

[47] Kyle Williams, Julia Kiseleva, Aidan C. Crook, Imed Zitouni, Ahmed Hassan Awadallah, and Madian Khabsa. 2016. Detecting Good Abandonment in Mobile Search. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. 495–505.

[48] Zhijing Wu, Mark Sanderson, Barla B Cambazoglu, Bruce W Croft, and Falk Scholer. 2020. Providing Direct Answers in Search Results: A Study of User Behavior. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '22)*. ACM, 1635 – 1644.

[49] Zhijing Wu, Mark Sanderson, B. Barla Cambazoglu, W. Bruce Croft, and Falk Scholer. 2020. Providing Direct Answers in Search Results: A Study of User Behavior. In *CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 1635–1644.

[50] Yusuke Yamamoto and Katsumi Tanaka. 2011. Enhancing Credibility Judgment of Web Search Results. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. 1235–1244.