Striving for Responsible Opinion Formation in Web Search on Debated Topics

Rieger, A.

DOI
10.4233/uuid:703a1aad-d585-459a-b0b3-ac55d9e98fcd

Publication date
2024

Document Version
Final published version

Citation (APA)
Rieger, A. (2024). *Striving for Responsible Opinion Formation in Web Search on Debated Topics*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:703a1aad-d585-459a-b0b3-ac55d9e98fcd

Striving for Responsible Opinion Formation in Web Search on Debated Topics

Striving for Responsible Opinion Formation in Web Search on Debated Topics

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Friday, 8th of November 2024 at 12:30 o'clock

by

Alisa Rieger

Master of Science in Sensors and Cognitive Psychology,
Chemnitz University of Technology, Germany
born in Erbach, Germany

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. G.J.P.M. Houben,	Delft University of Technology, *promotor*
Dr. M.S. Pera,	Delft University of Technology, *co-promotor*

Independent members:

Prof. dr. C.M. Jonker,	Delft University of Technology
Prof. dr. ir. D. Hiemstra,	Radboud University
Dr. J. Otterbacher,	Open University of Cyprus
Prof. dr. ir. I.R. van de Poel,	Delft University of Technology
Dr. A. Anand,	Delft University of Technology, *reserve member*

SIKS Dissertation Series No. 2024-25.
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Interactive *Natural Language* Technology for eXplainable Artificial Intelligence

NL4XAI

This project received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860621.

Keywords: Web Search, Debated Topics, Responsible Opinion Formation, Bias Mitigation, Searcher-System Interaction, Nudging, Boosting

Cover: Mara Süßmann (design) and Alisa Rieger (fish painting) – A school of fish, symbolizing the mutually reinforcing biases of searcher and system during web search on debated topics.

ISBN 978-94-6366-941-2

Contents

To everyone who consistently reminds me that life is so much more.

1

INTRODUCTION

In the years dedicated to writing this dissertation, we have witnessed a cascade of global crises, characterized by their staggering severity and ramifications, such as the COVID-19 pandemic, conflicts in Ukraine and Gaza, and the climate crisis, to name only a few. As with all turbulent periods, these years have also been marked by fierce debate, not only on how to respond to these crises but, more broadly, on how to restructure our societies and adjust the way we live in a changing world. The scope of these debates is broad, covering topics such as *sustainable lifestyle choices*, *reproductive rights*, *farming regulations*, *vaccine mandates*, or *police reforms*. Such *debated topics* are characterized by an ongoing discussion among groups of individuals with different perspectives and opinions, often tied to conflicting values or interests, and lacking a straightforward resolution [308]. Due to these tensions, interacting with different perspectives of debated topics can be cognitively demanding and elicit emotionally charged behavior and cognitive biases that hinder engagement with diverse viewpoints to gain a well-rounded understanding of the topic [124, 156, 269]. Simultaneously, discourse on such topics is crucial for a healthy democratic culture, wherein individuals actively engage with the various perspectives to *responsibly* form opinions, as these can shape practical decisions that affect their lives, communities, and the broader society [113, 129, 177, 227, 260].

Alongside fierce debates, over the past few years, we have also experienced a rise of right-wing populists, whose allure includes suggesting oversimplified answers and mobilizing collective emotions rather than critical thinking [92, 305]. This development has been suspected to be linked not only to the complex global crises we are witnessing but also to the changing information landscape and digital media uptake [134, 211, 258]. This is evidenced, for instance, by reports from 2022 which indicate that 84% of the EU population used the internet daily [86], with *finding information* being one of the main reported use-cases [65]. Yet, in the digital realm, access to information has largely been dictated by a small number of private companies, operating with limited public oversight, that tend to follow attention-driven profit models and leverage, for instance, choice architectures and

1

algorithmic content curation [199, 361]. Such structures might pose obstacles to exposure and engagement with diverse perspectives, as they, like right-wing populists, often rely on and optimize for emotional reactions rather than critical thinking [178, 210].

The Digital Humanism take. To better understand and mitigate hidden risks of digital technologies for individuals and democratic societies, proponents of the *Digital Humanism* initiative are advocating for thoroughly analyzing and reflecting on human-technology relationships, promoting the development of human-centered technology that prioritizes improved lives and societal progress over economic growth [371]. Doing so, however, is not an easy endeavor: humans and technology have co-evolved and mutually shaped one another, resulting in various hidden tensions that may surface unexpectedly [248, 380]. For instance, social media was envisioned to foster participation and democratic values, yet recent years have shown its capacity to contribute to a decline in political trust and democratic values [211, 243, 349].

Information behavior is changing. The advent of the internet has dramatically transformed our information behavior over the past decades, providing ubiquitous access to an ever-growing amount of online information [107, 178, 319]. While traditional media outlets often operate under regulations and ethical standards for a well-informed citizenry, e.g., regarding quality and diversity of the curated journalistic content [130, 240, 361], individuals increasingly rely on online environments to interact with information, which requires autonomous navigation of vast amounts of resources of varying quality and perspectives [134, 210, 342]. To effectively navigate the online world, search engines[1] that aim to ease the discovery of relevant documents, have become indispensable [48, 310].

Search engines have evolved. Since the early days of the web, search engines have adapted to the changing requirements and needs of different stakeholders. For instance, in the highly privatized digital realm, human attention has emerged as the primary driver of revenue, positioning search engines as algorithmic curators that have incentives to prioritize profit generation and individual user satisfaction over societal relevance and progress [48, 199, 214, 246, 330]. This is evident in practices like making use of implicit feedback, e.g., clicks and other user-system interaction behaviour, not only for enhanced relevance predictions and ranking functions [1, 89, 149], but also for strategic placement of sponsored content to increase users' click probability [57, 59, 133]. Additional developments like the widespread adoption of smartphones and the rise of social media have enabled nearly everyone to access, generate, and disseminate content on the web at any moment. This is why to ensure the retrieval of search results with high-quality information, search engine providers need to contemplate how to fairly assess and consider source credibility in

[1]Throughout this dissertation, the terms *search engines* and *search* are used interchangeably to refer to conventional *web search engines* and *web search*, respectively.

the search result ranking [51, 159, 197, 213, 273]. In response to such developments, search engines have become increasingly complex and opaque [115, 310, 320].

Searchers trust search engines. Despite their opacity, individuals have developed a strong sense of trust in search engines and often perceive them as value-neutral, and unbiased providers of information [115, 278]. Searchers tend to rely on the search engine's output and believe that it accurately and reliably compares, evaluates, and differentiates between different resources on their behalf, engaging primarily with high-ranked results [151, 320, 325]. Widely-used search engines like *Google* or *Bing* have incorporated functions such as providing direct answers to queries with *featured snippets*, or blending results from different resources into coherent text for conversational interfaces. Such trends towards providing desired information as seamlessly as possible further reinforce the tendency to rely on the search engine instead of actively exploring and comparing different resources [115, 147, 163, 320].

For instance, let us consider a searcher who wants to find information on *eel reproduction*. Depending on their query, they might be exposed to either of the Search Engine Result Pages (**SERPs**) shown in Figure 1.1. Regardless of whether the query is*where do baby eels come from* or *reproduction of eels*, the searcher is presented with information on eel reproduction. Yet, the featured snippets and search results highlight different aspects, resulting in slightly different learning outcomes, particularly when the searcher relies on the search engine and featured snippet rather than exploring lower-ranked search results.

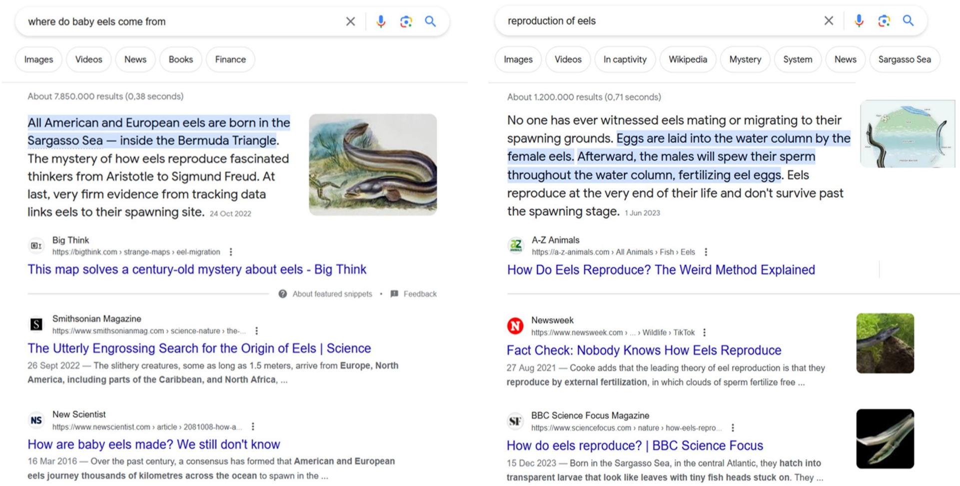


Figure 1.1: Example of two SERPs with featured snippets for different queries on *eel reproduction*, generated by Google in March 2024. For enhanced clarity, we excluded the *people also ask* suggestions displayed on the SERP.

Searchers' reliance on the search engine becomes problematic when they have complex and consequential information needs that inherently require the exploration of diverse resources and perspectives [226, 320, 365]. Nonetheless, people increasingly turn to search engines with such information needs that would require a comprehensive understanding of the information space to enable well-rounded knowledge gain, for instance, to form opinions and make practical decisions on debated topics [50, 96, 115, 370]. Search engines have thus evolved into crucial infrastructure for information access which wield considerable influence over individual perception, knowledge, decisions, and behavior [48, 105, 115, 117, 135].

Bias in search engines. Insights into how search engines accept, absorb, amplify, and reflect societal and internal human biases showcase that present-day search engines are far from neutral, raising questions on whether searchers' trust in them is warranted [22, 246, 372]. The influence of such biases on democratic societies becomes exceptionally salient in the context of political queries and elections, as search engine results were observed to shift voters' decisions when over-representing information about some candidates over others [82, 83, 266].

To better understand how search engines and their functions can (unintentionally) pose threats to democracy, researchers have conducted search engine audits, revealing imbalances in how search engines react to political queries in personalized but also non-personalized conditions [194, 303, 352]. This concern is not limited to the political realm. Similar imbalances have been observed by Draws et al. [72] for queries on debated topics, finding that even in response to neutral queries, search engines produced SERPs that over-represent one viewpoint over others (*exposure bias*) rather than providing diverse perspectives to enable well-rounded knowledge gain.

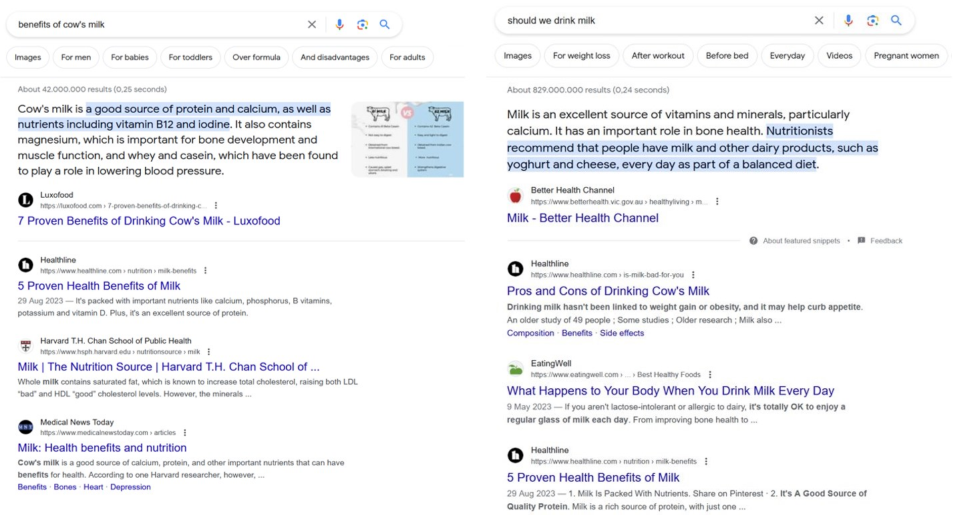


Figure 1.2: Example of two SERPs with featured snippets for different queries on *drinking cow's milk*, generated by Google in March 2024. For enhanced clarity, we excluded the *people also ask* suggestions displayed on the SERP.

Bias in searchers. Search on debated topics, however, is not only shaped by exposure biases but also by searchers' interaction choices [74, 219, 323]. These choices were found to reflect cognitive biases that are prevalent when engaging with debated topics, such as the tendency to prioritize information that aligns with searchers' preexisting attitudes while dismissing or discounting opposing information (*confirmation bias*) [18, 323, 360, 372]. Researchers have further observed an interplay of exposure and interaction biases [74, 82, 302, 323]. For instance, searchers were found to adopt the viewpoint that is over-represented on the SERP—a phenomenon known as the *search engine manipulation effect* [7, 82, 266]. Such insights underscore how the combined impact of exposure and interaction biases extends beyond the mere search process, shaping not only searchers' engagement with information and immediate knowledge gain, but potentially their opinions, perceptions, and practical decisions (e.g., whom to vote for [82]) with considerable consequences for both individuals and society.

The searcher-system dynamics in practice. The entangled searcher-system dynamic marked by exposure and interaction bias, along with its consequential nature, distinguishes search on debated topics from other search tasks, like searching for information on *eel reproduction* (see Figure 1.1). To illustrate, let us consider two individuals exploring information about consuming cow's milk: Jo, raised on a dairy farm and following a diet that is primarily based on animal products, and Max, contemplating switching to a vegan diet for sustainability reasons but worried about nutritional deficits. Jo queries for *benefits of cow's milk* and is exposed to the SERP depicted on the left of Figure 1.2. Being strongly opinionated and having a sense of expertise due to their upbringing, they would find their beliefs reinforced by the snippet and search results. Conversely, Max, who does not have a strong opinion but an interest in gaining knowledge to make an informed decision formulates the more neutral query, *should we drink milk*, and gets exposed to the right SERP in Figure 1.2. While the snippet and top-ranked search results mostly emphasize the benefits of drinking milk, they might be open to exploring beyond the high-ranked search results to engage with diverse viewpoints. However, having learned to rely on search engines providing the most relevant information on top, they might be discouraged from switching to a vegan diet.

Towards responsible opinion formation in web search. Recent research has suggested visions of future online environments and search systems, designed to support users in overcoming the inherent challenges of navigating the web and undertaking complex information-seeking endeavors [178, 210, 320]. Inspired by these proposals, we posit that search engines have the potential to be platforms that encourage and empower informed and autonomous choices, supporting the exploration of debated topics in their full complexity. However, the challenges and risks associated with web search on debated topics can only be understood and mitigated from a perspective that centers around searchers and their interaction choices [219, 302, 323]. To this end, this dissertation takes a searcher-centered inventory of the distinct challenges of the searcher-system interplay and potential

1

remedies to address them. Our primary focus lies in understanding the factors inherent to both the searcher and the search system that influence interaction choices. Through this approach, we aim to gain insights into factors that either facilitate or hinder unbiased and diligent engagement with diverse viewpoints and, in consequence, fruitful search on debated topics. These insights can provide valuable guidance towards designing search engines and interfaces that support searchers in overcoming the challenges of search on debated topics, ultimately enabling responsible opinion formation.

1.1. Research Questions and Contributions

This dissertation is divided into four parts, each advancing the journey to enable responsible opinion formation in web search on debated topics by focusing on a distinct aspect: understanding the searcher-system interplay (Part I), guiding search behavior (Part II), empowering searchers (Part III), and broader research challenges towards responsible opinion formation (Part IV). In the remainder of this section, we describe for each of the four parts the underlying motivation, overarching research question, applied methods, and derived insights and contributions. We explain and define important terms when they are first introduced, highlighting each term in italics. Further, we repeat these explanations and definitions in the relevant parts. For easy access, we also provide a glossary with the key terminology (Page 163).

Part I: Understanding the Searcher-System Interplay

Search engines retrieve, rank, and present search results to the searcher whereas searchers interact with the search engine, for instance by querying, scanning the results, and clicking on selected results to navigate to the linked web pages, which might cause changes in their knowledge and attitude on the topic. Search on debated topics is characterized by attributes inherent to the searcher, the search system, and their mutually evolved and convoluted interplay. For instance, biases in the search engine ranking (e.g., *exposure bias*—one viewpoint is over-represented among high-ranked search results) and in searchers' interaction choices (e.g., *confirmation bias*—searchers primarily click on attitude-confirming search results, or *position bias*—searchers primarily click on high-ranked search results) collectively impact search behavior [323] and searcher's *epistemic states*—user states (i.e., temporary conditions) that are related to knowledge and opinions, such as their attitude change [74, 219]. To enable responsible opinion formation in web search on debated topics, it is crucial to first *understand* how various factors and biases inherent to the searcher and the search system, individually and collectively, shape web search on debated topics. Therefore, with Part I, we address the following research question:

> **RQ_{I}**: How do characteristics of the searcher and search system shape search on debated topics?

To shed light on the complex dynamics of the searcher-system interplay, we addressed **RQ**$_{\mathbf{I}}$ with two exploratory user studies, capturing a wide range of variables to gain comprehensive initial insights [315].

We first approached **RQ**$_{\mathbf{I}}$ by studying the relations between (i) factors inherent to the searcher (attitude strength, prior knowledge, *receptiveness to opposing views*—the willingness to access, consider, and evaluate opposing views in an impartial manner [229]) and search system (exposure bias), (ii) search interaction (confirmation bias, position bias, search effort), and (iii) post-search *epistemic states* (attitude change, knowledge gain) with an exploratory, open-ended user study with 255 participants, which we discuss in Chapter 2. Our findings suggest that search interactions were shaped by exposure bias, as well as searchers' attitude strength, and prior knowledge. Moreover, attitude change was affected by confirmation bias and initial attitude strength. These findings imply that pre-search epistemic states play a primary role in shaping search on debated topics. Specifically, searchers with strong prior attitudes exhibited substantial confirmation bias and low knowledge gain and were highly unlikely to change their attitudes.

To deepen the investigation into searchers' reliance on the search ranking, we center Chapter 3 around the exploration we conducted to gain a better understanding of the effects of exposure bias by investigating conditions of increased ranking transparency. For that we conducted a user study with $N = 198$ participants, testing the effect of stance labels that indicate the stances of the search results (i.e., *neutral, opposing, supporting*) on search interactions and attitude change. Our findings indicate that stance labels increase viewpoint diversity in search interactions and mitigate the effects of ranking bias, thus fostering more appropriate reliance on the search system and ranking.

Contributions of Part I

- We present findings of an exploratory user study with 255 and a preregistered user study with 198 participants, investigating the interplay of searcher-rooted interaction and system-rooted exposure biases and the role of ranking transparency in shaping search on debated topics.

- We make the data sets with interaction data and questionnaire responses publicly available to allow fellow researchers to replicate, repurpose, and build on our data and findings.

- We highlight the risk of high user reliance on opaque search engine rankings and the amplification of user biases through personalized search result rankings.

- We showcase the effectiveness of stance labels to increase the transparency of the viewpoint ranking in mitigating the effects of exposure bias, promoting appropriate reliance on the search ranking.

- We publish validated knowledge questionnaires to measure individuals' knowledge on *abortion* and *obesity* which can be leveraged by the research community.

1

Part II: Guiding Search Behavior

In light of the noteworthy behavioral patterns of strongly opinionated searchers (e.g., high confirmation bias, low knowledge gain), uncovered as a result of the work detailed in Chapter 2, Part II centers on mitigating confirmation bias of strongly opinionated searchers. For that, nudging interventions, aimed at *guiding search behavior* towards decreased interactions with attitude-confirming search results emerged as a useful resource [49, 148, 337]. The concept of *nudging* refers to subtly guiding individuals to make decisions that are considered to be beneficial for them, without restricting possible choices [49, 337]. Inspired by nudges that successfully reduce engagement with misinformation [55, 157, 179, 221], we turned towards *warning labels* and *obfuscations* to flag and decrease the ease of access to attitude-confirming search results. Yet, although generally successful in guiding user behavior, nudging strategies run the risk of harming user autonomy [49, 119, 210]

Therefore, in Part II, we focus on the following research question:

> **RQ$_{\text{II}}$**: Can we guide individuals with warning labels and obfuscations to engage in unbiased search behavior on debated topics without harming their autonomy?

To answer this **RQ$_{\text{II}}$** we conducted two user studies. With the first study, we investigated the effect of warning labels with obfuscations on searchers' clicks on attitude-confirming search results with a preregistered user study with 282 participants that we present in Chapter 4. To identify whether the intervention could be misused to steer search interactions for purposes other than confirmation bias mitigation, we investigated not only the effect of warning labels with obfuscations applied to attitude-confirming but also to randomly selected search results. We found that the intervention reduced interactions with targeted search results, independent of whether they were attitude-confirming or randomly selected.

From the results emerging from this first study, it was unclear whether the warning label, the obfuscation, or both caused the reduced interaction. Further, exploratory findings indicated that the extent to which the warning label and the obfuscation were the sources of decreased interaction might vary across users with distinct cognitive styles. Thus, we conducted a follow-up study, presented in Chapter 5, to better understand which element of the intervention (warning label or obfuscation) caused the effect of decreased interaction and the role of searchers' cognitive style. In this second study with 307 participants, we found that obfuscations harm user autonomy and run the risk of manipulating instead of guiding searchers, while the warning label without obfuscation still motivated searchers to actively choose to decrease engagement with attitude-confirming search results.

Contributions of Part II

- We present findings of two preregistered user studies with 282 and 307 participants, investigating the effect of warning labels and obfuscations on searchers' confirmation bias, revealing benefits and risks of warning labels and obfuscations to mitigate confirmation bias among diverse users.
- We make the material (e.g., 200 viewpoint-annotated search results on four topics) and data set with interaction data and questionnaire responses from both user studies publicly available to allow fellow researchers to replicate, repurpose, and build on our data and findings.
- We validate the findings on the effect of warning labels with obfuscations on confirmation bias presented in Chapter 4 through replication in Chapter 5.
- We give an account of the ethical implications of automatic nudging interventions to guide individuals' information behavior, highlighting specific risks.
- We provide design guidelines for nudging interventions that aim at supporting unbiased search behavior during web search as a foundation for future research into such interventions.

Part III: Empowering the Searcher

Reflecting on the complex dependencies on individuals' epistemic states that we observed in Part I, we came to realize that merely steering searchers towards reduced interaction with attitude-confirming search results, i.e., mitigating confirmation bias, does not adequately support fruitful search on debated topics. This requires unbiased and also diligent search behavior, such as actively gathering, evaluating, and comparing information on various arguments to gain a well-rounded understanding of the topic [177, 260, 320].

In an attempt to overcome some of the limitations of nudging interventions, namely the risk of harming user autonomy and failure to address the challenges related to search on debated topics more comprehensively and sustainably, we hence shifted the focus from directly *guiding search behavior* towards *empowering the searcher* to overcome the challenges of web search on debated topics with autonomy-preserving boosting interventions. Boosting interventions seek to foster lasting user competencies to navigate different challenges and should remain effective after being presented to the individual [132, 210]. To boost searchers' competencies in navigating search on debate topics, we target their metacognitive skills, specifically their *intellectual humility* **(IH)** which encompasses the competencies to recognize the limits of one's knowledge and be aware of the fallibility of one's opinions and beliefs [267]. Part III hence evolves around the following research question:

> **RQ$_{\text{III}}$**: Can we empower individuals to engage in unbiased, as well as diligent search behavior, with interventions that boost their intellectual humility?

To address **RQ**$_{\text{III}}$, we examine cognitive biases, specifically confirmation bias, in the information search process on debated topics and contrast nudging and boosting interventions based on a review of the literature. Motivated by insights on the potential of boosting interventions to mitigate confirmation bias in addition to more broadly encouraging diligent search on debated topics, we propose a vision to shift towards interventions that boost searchers' IH in Chapter 6. To initiate the pursuit of this vision, we outline research questions and potential research challenges.

Following up on this vision, we identified three interventions that boost self-reported IH in a pre-study. We tested their effect on opinionated individuals' search behavior in a familiar search environment and explored the role of IH in the search process more broadly by conducting a preregistered user study with 299 participants which we present in Chapter 7. We did not find effects on strongly opinionated searchers' behavior. Yet, explorations revealed that searchers' level of intellectual humility is relevant for cultivating rewarding search experiences but might require a more direct integration in the search process and combination with additional strategies to foster unbiased and diligent search behavior, for instance increasing the viewpoint-ranking transparency (see Part I).

Contributions of Part III

- We propose and motivate a vision of shifting from guiding search behavior to empowering searchers to overcome the challenges associated with search on debated topics, including a set of initial research questions and potential research challenges.

- We design three interventions to boost IH and demonstrate their effectiveness in boosting self-reported IH in a pre-study.

- We present a preregistered user study with 299 participants, testing the potential of IH boosts in practice by investigating the effect of the three interventions that boost self-reported IH on individuals' search behavior in a familiar search environment and exploring the broader role of IH in the search process.

- We make the data collected with the user study (interaction data and questionnaire responses) publicly available to allow fellow researchers to replicate, repurpose, and build on our data and findings.

- We outline design implications for boosting interventions aimed at supporting unbiased and diligent search behavior.

- We discuss methodological insights for evaluating interventions that are aimed at modifying online information behavior.

Part IV: Towards Supporting Responsible Opinion Formation

To this point, the entire dissertation journey was marked by unveiling additional layers of complexity inherent to web search on debated topics at every step. However, as we investigate issues that ultimately concern information access and societal well-being, it is essential to acknowledge and navigate this complexity, resisting the temptation of simple fixes, but instead ensuring that solutions truly mitigate rather than inadvertently exacerbate harm [91, 243]. Reflecting on our findings, we want to emphasize that in striving to enable responsible opinion formation in web search, we stand merely at the beginning. In Part IV we hence target the following research question:

> **RQ$_{\text{IV}}$**: What are challenges and research opportunities towards supporting web search on debated topics, promoting responsible opinion formation?

In Chapter 8 we address **RQ$_{\text{IV}}$** by outlining challenges and research opportunities inherent to the searcher, the search system, and their interplay, building on perspectives from digital humanism and through an extensive interdisciplinary literature review, to provide a foundation for future research efforts in the realm of web search on debated topics.

Contributions of Part IV

- We give an account of the high-level challenges of web search on debated topics, building on perspectives from digital humanism.
- We present a review of interdisciplinary literature on the challenges associated with the searcher, the search system, and their interplay during web search on debated topics.
- We provide a research agenda encompassing high-level challenges, methodological considerations, and concrete initial research questions for research endeavors concerning web search on debated topics.

1.2. Positioning and Impact

With the research presented in the four parts of this dissertation, we advance knowledge in the fields of human information interaction, behavioral change interventions, and user modeling and adaptation. We demonstrate the distinct characteristics of web search on debated topics, emerging from the entangled and consequential interplay of searchers' interaction choices and search system exposure. In consequence, we underscore the need for dedicated research efforts. Our insights on the benefits and risks of interventions to guide and empower unbiased and diligent search on debated topics, as well as methods to investigate such approaches, establish a basis for designing, researching, and personalizing additional interventions for search environments that support searchers. We provide

1

a foundation for future research by giving a detailed account of the challenges on the path towards responsible opinion formation in web search and outlining potential approaches to tackle these challenges.

Before moving on to the chapters where we describe the research we have conducted within the scope of this dissertation, we want to emphasize Heisenberg's cautionary words that apply to both conducting research and forming opinions on debated topics: *We have to remember that what we observe is not nature herself, but nature exposed to our method of questioning* [127].

1.3. DISSERTATION ORIGINS

The four parts of this dissertation are based on the publications listed below. These publications are the result of collaboration between my co-authors and me. Modifications to the publications were kept to a minimum, focusing on reconciling terms, unifying the structure, reducing repetitiveness, and incorporating feedback from the doctoral committee. We apply a local numbering of research questions within each chapter.

- A full conference paper, published at the ACM conference on User Modeling, Adaptation and Personalization (UMAP) 2024: **Alisa Rieger**, Suleiman Kulane, Ujwal Gadiraju, and Maria Soledad Pera. "Disentangling Web Search on Debated Topics: A User-Centered Exploration" [299] (Part I)

- A full conference paper, published at the ACM conference on Hypertext and Social Media (HT) 2021: **Alisa Rieger**, Tim Draws, Mariët Theune, and Nava Tintarev. "This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias" [298] 🏆[2] (Part II)

- A journal paper, published in the ACM journal on Transactions on the Web (TWEB) 2023: **Alisa Rieger**, Tim Draws, Mariët Theune, and Nava Tintarev. "Nudges to Mitigate Confirmation Bias during Web Search on Debated Topics: Support vs. Manipulation" [297] (Part II)

- A late-breaking paper, published at the ACM conference on Human Factors in Computing Systems (CHI) 2023: **Alisa Rieger**, Frank Bredius, Nava Tintarev, and Maria Soledad Pera. "Searching for the Whole Truth: Harnessing the Power of Intellectual Humility to Boost Better Search on Debated Topics" [295] (Part III)

- A full conference paper, published at the ACM conference on Human Information Interaction and Retrieval (CHIIR) 2024: **Alisa Rieger**, Frank Bredius, Mariët Theune, and Maria Soledad Pera. "From Potential to Practice: Intellectual Humility During Search on Debated Topics" [294] (Part III)

[2]The trophy icon indicates that the paper won a best paper award

- A full conference paper, published at the European Conference on Information Retrieval (ECIR) 2024: **Alisa Rieger**, Tim Draws, Nicolas Mattis, David Maxwell, David Elsweiler, Ujwal Gadiraju, Dana McKay, Alessandro Bozzon, and Maria Soledad Pera. "Responsible Opinion Formation on Debated Topics in Web Search" [296][3] (Part IV)

This dissertation benefited from the following additional conference, workshop, and doctoral consortium publications that advanced our understanding of the challenges of web search on debated topics and human bias mitigation, focusing on cognitive bias mitigation in online interactions other than web search (e.g., online debates and crowd work), and viewpoint diversity in search results:

- Tim Draws, Nirmal Roy, Oana Inel, **Alisa Rieger**, Rishav Hada, Mehmet Orcun Yalcin, Benjamin Timmermans, and Nava Tintarev. "Viewpoint Diversity in Search Results" [72]
- Zhangyi Wu, Tim Draws, Federico Cau, Francesco Barile, **Alisa Rieger**, and Nava Tintarev. "Explaining Search Result Stances to Opinionated People" [382]
- **Alisa Rieger**, Qurat-Ul-Ain Shaheen, Carles Sierra, Mariet Theune, and Nava Tintarev. "Towards Healthy Engagement with Online Debates: An Investigation of Debate Summaries and Personalized Persuasive Suggestions" [300]
- **Alisa Rieger**. "Interactive Interventions to Mitigate Cognitive Bias" [293]
- Tim Draws, **Alisa Rieger**, Oana Inel, Ujwal Gadiraju, and Nava Tintarev. "A Checklist to Combat Cognitive Biases in Crowdsourcing" [71]
- **Alisa Rieger**, Mariët Theune, and Nava Tintarev. "Toward Natural Language Mitigation Strategies for Cognitive Biases in Recommender Systems" [301]

[3]The first two authors made equal contributions to this publication

I

Understanding The Searcher-System Interplay

Search on debated topics is characterized by attributes inherent to the searcher, the search system, and their mutually evolved interplay. For instance, search engines can produce *exposure bias*, i.e., search engine result pages (**SERPs**) on which one viewpoint is over-represented [72]. Further, searchers' cognitive biases can manifest in their search interactions, for instance, when they favor attitude-confirming over attitude-opposing information (*confirmation bias*) or resort to high-ranked search results (*position bias*) [18]. Observations such as the *search engine manipulation effect*, searchers shifting their attitude towards the viewpoint over-represented on the SERP [7, 82, 266], showcase that exposure and interaction biases can interact. Yet, searchers have learned to rely on opaque search engine rankings, perceiving them as neutral and unbiased providers of information, unaware of the impact that exposure and interaction biases have on their search experience, opinions, and decisions [115, 278, 325]. In light of this searcher-system interplay, it is essential to gain a comprehensive understanding of how various factors and biases inherent to the searcher and the search system, individually and collectively, impact web search on debated topics. Therefore, we pose the following research question:

RQ$_{\mathrm{I}}$: How do characteristics of the searcher and search system shape search on debated topics?

Part I consists of two chapters, each presenting a user study to answer **RQ$_{\mathrm{I}}$**. In both studies, we emphasized exploring a wide range of different variables to gain comprehensive initial insights [315]. In Chapter 2 we present the first user study, in which we investigated the relations between (i) factors inherent to the searcher (attitude strength, prior knowledge, receptiveness to opposing views) and search system (exposure bias), (ii) search interaction (confirmation bias, position bias, search effort), and (iii) post-search *epistemic states*—user states that are related to knowledge and opinions (attitude change, knowledge gain). Our findings imply that in addition to exposure bias, pre-search epistemic states play a primary role in shaping search on debated topics. Specifically, searchers with strong prior attitudes exhibited high confirmation bias and low knowledge gain and were unlikely to change their attitudes. In Chapter 3, we present a user study, aimed at gaining a better understanding of the effects of exposure bias by investigating conditions of increased ranking transparency. For that, we tested the effect of *stance labels*—labels that show the stance of a search result on a specific debated topic (i.e., *neutral, opposing, supporting*) on search interactions and attitude change. Our findings indicate that stance labels increase viewpoint diversity in search interactions and mitigate the effects of ranking bias. These findings highlight that searchers will detect exposure bias and engage with diverse viewpoints if provided with increased viewpoint-ranking transparency.

2

Disentangling Web Search on Debated Topics: A User-Centered Exploration

This chapter is based on a full conference paper: **Alisa Rieger**, Suleiman Kulane, Ujwal Gadiraju, and Maria Soledad Pera. "Disentangling Web Search on Debated Topics: A User-Centered Exploration". In: *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 24–35. ISBN: 9798400704338. DOI: 10.1145/3627043.3659559

2.1. Introduction

The body of work dedicated to understanding the impact of exposure and interaction biases during web search on debated topics, along with the strategies to overcome the associated challenges, is growing. Research, however, has predominantly investigated specific aspects of this larger issue in isolation. For instance, recent works studied relations between exposure and interaction [298, 382], user traits and interaction [294, 297], and exposure, interaction, and attitude change [74, 173, 323], frequently focusing specifically on individuals with either strong or weak pre-search attitudes. Across these works, a recurring point of discussion is the difficulty of understanding the intricate searcher-system interplay that is characterized by entangled effects of system exposure, search interactions, and user characteristics. We argue that comprehension of the complex dynamics of web search on debated topics requires undertaking a holistic approach with the user placed at the center of this exploration, as depicted in Figure 2.1.

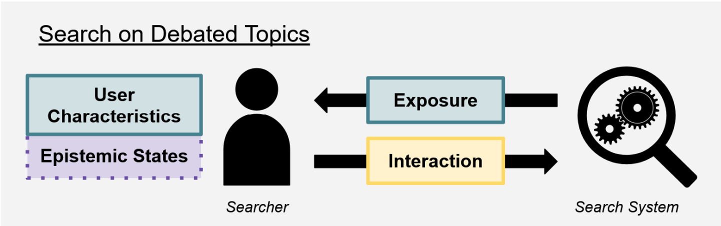


Figure 2.1: **A *user-centered perspective* of search on debated topics.** Searchers with varied characteristics are *exposed* to resources on debated topics by a search system. They choose the results to *interact* with, modifying their epistemic states.

To advance knowledge and uncover relations between the different facets of web search on debated topics, we adopted an exploratory, open-ended approach. We studied the relations between factors inherent to the *searcher and search system* (user characteristics, exposure bias), *search interaction* (confirmation bias, position bias, search effort), and *post-search epistemic states* (attitude change, knowledge gain), as illustrated in Figure 2.2. This is inspired by outcomes inferred from research on belief dynamics during web search for information on medical yes-no questions with a known true answer (expert consensus) [373, 376]. In this context, White [373] observed that pre-search beliefs affected interaction and that exposure bias could shift post-search beliefs, but only if participants did not have strong pre-search beliefs. However, the investigated yes-no questions are considerably different from debated topics that involve multiple viewpoints and have no defined correct answer. In addition to exposure bias and beliefs, users' prior knowledge and personality traits might play a role in shaping search interaction and post-search epistemic states [297]. These are recognized factors influencing general web search behavior [151, 205, 309], yet, their combined impact in the context of debated topics remains to be explored.

In this study, we investigated the following research questions:

RQ1 How do attributes of the searcher and search system shape search interaction?

RQ2 How do attributes of the searcher and search system shape the post-search epistemic states of the searchers?

RQ3 How do search interactions shape the post-search epistemic states of the searchers?

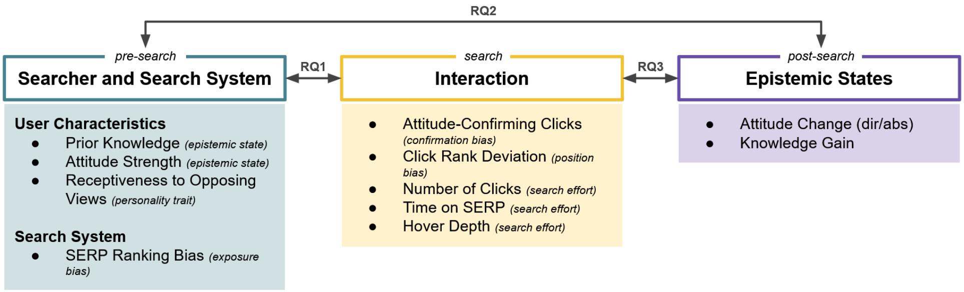


Figure 2.2: **Exploration Overview.** To gain a better understanding of web search on debated topics, we investigated RQ1, RQ2, and RQ3, exploring relations between selected attributes of the searcher and search system, the search interaction, and the post-search epistemic states.

To address these RQs, we conducted a user study with 255 participants and asked them to imagine a scenario where they would use a search system to find arguments on a debated topic. Before the search task, we captured user characteristics, i.e., participants' epistemic states and the relatively stable trait of receptiveness to opposing views [229]. During the search task, we exposed participants to a mock SERP with results portraying different viewpoints (i.e., *opposing, supporting, neutral*) towards the topic statement. We logged their interactions to then compute metrics that approximate their confirmation bias, position bias, and search effort. Using a between-subjects experimental design, we manipulated exposure by assigning participants to three SERP ranking bias conditions: balanced, biased supporting, and biased opposing. We captured their post-search epistemic states(attitude change and knowledge gain).

The findings of this empirical study exploring relations among attributes of the searcher and search system, search interaction, and user post-search epistemic states in the context of web search on debated topics indicate that search interaction was shaped by search system exposure, attitude strength, and prior knowledge. Attitude change was not directly affected by search system exposure but by participants' confirmation bias and initial attitude strength. We observed various moderating effects, which suggest that effects of exposure, interaction, and their interplay are moderated by prior knowledge, attitude strength, and potentially other pre-search epistemic states. These observations can serve as a foundation for research aiming to identify strategies to overcome the risks related to web search on debated topics, particularly for adapting such strategies to different searchers' needs. The dataset, containing behavioral data from search logs and measures of knowledge, attitude, receptiveness to opposing views from questionnaire responses, as well as the material used for data collection are publicly available.[1]

2.2. Related Work

Engaging with Debated Topics. Web search engines are used to fulfill wide-ranging information needs, including complex inquiries that require exploration, such as researching information about debated topics [15, 50, 97, 377]. Engaging with information on debated topics can cause changes in individuals' opinions and knowledge, thus their *epistemic states* [82, 173]. *Epistemics* is an interdisciplinary field focusing on information processing, knowledge behavior, and belief formation which incorporates insights from epistemology, cognitive science, and information science, among other disciplines [164]. We refer to user states that are related to knowledge and opinions as user or searcher *epistemic states*. Opinions on debated topics can lead to practical decisions, e.g., whether to adopt vegetarianism, equally split parental leave, or vote for a certain party. Given the substantial implications of such decisions for individuals and society, searchers would ideally engage with diverse viewpoints to become informed on the topic, enabling them to form opinions *responsibly* [177, 260, 296, 297].

Interaction and Exposure Biases. For the individual, interacting with information on debated topics can be challenging and affected by cognitive biases, shaping their interactions with the search system and impeding responsible opinion formation [18, 173, 296]. When searching about debated topics, users may, for instance, prioritize protecting and defending their own beliefs and values over the pursuit of becoming informed by gaining knowledge about diverse viewpoints [124, 140, 156, 269], thus interacting preferably with information that aligns with preexisting beliefs (*confirmation bias*) [18, 360, 372]. Further, searchers have learned to rely on search engines and mostly interact with highly ranked search results (*position bias*) [151] trusting the search engine to provide relevant, unbiased, and credible information [115, 122, 278, 325]. Yet, recent research by Draws et al. [72] has observed viewpoint-biased rankings in response to queries on debated topics across different search engines, characterized by an over-representation of specific viewpoints in highly ranked search results, even in response to viewpoint-neutral queries. Search engines might further amplify searchers' interaction biases by tailoring the search result ranking to maximize individual relevance, for instance, based on click rates [22, 48, 115, 330]. Such biases commonly remain unnoticed by searchers who, due to the opacity of automatic filtering and ranking processes, face difficulties in determining whether the provided search results are unbiased and complete [226, 325]. Conversely, searchers were even observed to adopt the prevailing viewpoint when exposed to SERPs with a viewpoint-biased ranking—a phenomenon known as the *search engine manipulation effect* [82]. For search on debate topics, however, Draws et al. [74] did not find evidence for an effect of exposure bias on attitude change, but instead found that it was linked to search interactions. For search on political topics and news, studies indicate that although exposure plays a considerable role in shaping interactions, individuals still tailor their

[1] https://osf.io/u3s5n/?view_only=86cb2495551943bd87576d9790aef3dd

interactions to align with prior beliefs when faced with belief inconsistent exposure bias [302, 323].

User Characteristics. The interaction with information on debated topics can vary based on user characteristics, such as their epistemic states and more stable user traits. For instance, the manifestation of confirmation bias in an individual's information behavior was found to be influenced by their attitude strength, where those with weaker attitudes seem to be more likely to engage with and open to processing attitude-opposing information [124, 172, 358]. Further, prior research has observed variance in web search behavior depending on users' topic knowledge [93, 205, 374, 389]. Searchers with high topic knowledge tend to employ more efficient search strategies [374], demonstrating a reduced susceptibility to position bias by being more likely to click on lower-ranked items on the SERP [205] and select items based on topic relevance and source credibility [308]. In the context of search on debated topics, there are only preliminary explorations on the role of prior knowledge [297]; more conclusive insights are pending.

In addition to epistemic states, relatively stable user traits are known to shape interactions with information on debated topics [47, 121, 195, 268]. Particularly relevant in the context of debated topics and responsible opinion formation is the trait of *receptiveness to opposing views*–defined as the *willingness to access, consider, and evaluate opposing views in an impartial manner* [229].

2.3. METHOD

We conducted and exploratory between-subjects study to investigate our research questions. All related material can be found at the URL in Footnote 1.

2.3.1. EXPERIMENTAL SETUP

To probe the dynamics of web search on debated topics, we selected two topics from *ProCon* [275], a resource that presents controversial topics and related arguments, with varying levels of controversy: '*Should abortion be legal?*' (highly controversial, people tend to have strong attitudes) and '*Is obesity a disease?*' (moderately controversial, people tend to have moderate attitudes). We created a custom SERP on which we displayed ten ranked, pre-selected, viewpoint-annotated search results on a given topic. As in conventional SERPs, we displayed the title and snippet for each SERP result. By clicking on a result searchers could access the corresponding linked webpage.

To manipulate the exposure bias, we assigned participants to one of three SERP ranking bias conditions, in which the results were ranked adhering to a viewpoint ranking template: *balanced*, *biased supporting*, and *biased opposing*. In the *balanced* condition, participants were exposed to alternating attitude-confirming and -opposing results. We randomly varied whether an opposing or confirming result would be displayed on the first rank. In the *biased* ranking conditions, the first six were either attitude-confirming (biased supporting) or attitude-opposing (biased opposing) search results, followed by two neutral and two attitude-opposing or attitude-confirming ones, respectively. By including two neutral search results we could derive additional insights into the interplay of exposure and interaction bias by assessing whether participants would invest the extra effort to interact with the two lowest-ranked search results to learn about the underrepresented viewpoint or engage with search results that confirm their opinion.

We selected the ten SERP results per condition from a set of viewpoint-annotated search results that we prepared for each topic. For this, we obtained 30 search results per topic from the *Bing API* [223] which had to fulfill our inclusion criteria (no paywall, content focuses clearly on the topic). We collected annotations (on a 7-point Likert scale, ranging from *strongly opposes* to *strongly supports* the topic statement) from crowd workers recruited via *Prolific* [276]. Each crowd worker annotated ten search results, and each search result was annotated by three crowd workers. To control for quality, we included two attention checks and discarded annotations from crowd workers who failed at least one. Ultimately, we assigned the median value of the three annotations to the search results.

To measure prior knowledge and knowledge gain, we developed knowledge questionnaires. Per topic, we compiled a list of 60-80 statements (technically referred to as items) based on details extracted from the *Wikipedia* pages on the topic (e.g., *Approximately 45% of abortions conducted globally are considered unsafe, According to the American Medical Association, obesity is a disease*). We recruited 20 distinct crowd workers per topic to judge these items and respond with *true*, *false*, or *I don't know* labels. We discarded items that proved to be too easy (resulting in all correct responses) or too difficult/ambiguous (resulting in all incorrect and *I don't know* responses). We then created the knowledge questionnaires by selecting a subset of items per topic that could capture participants' knowledge reliably [93]. To maximize the internal reliability of the knowledge questionnaires we randomly selected 4,000 samples of 15 items from the remaining item pools per topic and computed Cronbach's α for each sample. We identified a set of 15 items per topic with Cronbach's α exceeding 0.8. The knowledge questionnaires used in our study are publicly available at the link in Footnote 1.

To measure participants' prior **attitude** and attitude change, we asked them to report their agreement with a statement on the assigned topic (i.e., '*obesity is a disease*', '*abortion should be legal*') on a 7-point Likert scale ranging from *strongly disagree (-3)* to *strongly agree (+3)* before and after the search task, adopting the approach used in prior research on attitude change in web search (e.g., [74, 82, 297]). We measured participants' **receptiveness to opposing views** with the 18-item self-report measure developed by Minson et al. [229] which captures *negative emotional reactions toward disagreement, intellectual curiosity regarding opposing views, derogation of those holding opposing views*, and *belief that it is inappropriate to debate certain issues.*

2.3.2. PROCEDURE

We recruited participants via *Prolific* [276] and paid them 2.1£ (mean across participants = 7.90£/h) for their participation, following the Prolific recommendations for fair pay at the time of data collection. They had to be at least 18 years old and proficient in English. The questionnaire responses were collected using

Table 2.1: **Study Variables.** Name, values, and description of the variables that we manipulated or measured. For searcher and search system variables we used two representations to compute correlations (numerical) and investigate group differences (categorical).

	Name	Values	Description
Pre-Search Searchers and Search System	SERP Ranking Bias	balanced (0), biased supporting (1), biased opposing (-1)	Randomly assigned to each participant
	Prior Knowledge	0 to 1 (correlations)	The Proportion of correctly answered questions of the knowledge questionnaire
		low, moderate, high (group differences)	Derived from the distribution of knowledge scores across all participants, those in the lowest quartile are categorized as having low knowledge; those in the highest quartile, high knowledge
	Attitude Strength	undecided (0), weak (1), moderate (2), strong (3)	Reporting to neither agree nor disagree with a topic statement was considered as an *undecided*, to somewhat agree or disagree as a *weak*, to agree or disagree as a *moderate*, and to strongly agree or disagree as *strong* attitude
	Receptiveness to Opposing Views	-1 to 1 (correlations)	Higher values indicate a higher receptiveness
		low, moderate, high (group differences)	Derived from the distribution of receptiveness to opposing views scores across all participants, those in the lowest quartile are categorized as having a low receptiveness, and the highest quartile as having a high receptiveness to opposing views
Search Interaction	Attitude-Confirming Clicks	0 to 1	The proportion of attitude-confirming results among the search results participants clicked on *(only for participants who clicked on one or more search results)*
	Click Rank Deviation	0 to 1	Deviation of the mean rank clicked from the mean rank if the participant would have clicked the top-ranked search results, normalized by the number of clicks *(only for participants who clicked on one or more search results)*; A value of 1 indicates maximal deviation (the participant clicked on the lowest-ranked search results), a value of 0 indicates no deviation (the participant clicked the top-ranked search results)
	Number of Clicks	0 to 10	The number of distinct search results a participant clicked on
	Time on SERP	in seconds	Amount of time that a participant spent on the search task in seconds
	Hover Depth	0 to 10	Lowest ranked search result a participant hovered on
Post-Search Epistemic States	Attitude Change	-6 to 2 (Directional)	The difference between the pre- and the post-search attitude; Negative values indicate an attitude weakening or change to the opposing attitude, while positive values indicate an attitude strengthening. Since directional attitude change could only be calculated for participants who did not report to be *undecided*, the values could range between -6 (change from *strongly agree* to *strongly disagree* or vice versa) and +2 (change from *somewhat agree* to *strongly agree* or *somewhat disagree* to *strongly disagree*)
		0 to 6 (Absolute)	The absolute difference between the pre- and the post-search attitude
	Knowledge Gain	-1 to 1	The difference between pre and the post-search knowledge; Negative values indicate a loss and positive values a gain of knowledge

Qualtrics [283]. We integrated five attention checks across the pre and post-search questionnaires, in which we instructed participants on which response to select. All data was collected in January 2023 with the following procedure, approved by our institution's ethics committee:

- **Pre-Search:** Given the potential sensitivity of the debated topics (i.e., abortion, obesity), we named them in the opening statement to allow prospective participants to make an informed decision regarding their participation. After receiving participants' informed consent to join the study, we randomly assigned them to one of the two debated topics and asked them to state their attitude on the topic (**attitude strength**). We then assigned them to one of the three **SERP ranking bias** conditions, balancing the distribution of participants with different attitude strengths across conditions. Subsequently, we asked them to fill out the knowledge questionnaire (**prior knowledge**) on the assigned topic and measured their **receptiveness to opposing views**.
- **Search:** For the search task, we asked participants to envision a situation in which they prepared for a mock debate with colleagues by making use of the search engine to find arguments related to the assigned topic. They could access the custom search page and click on search results to retrieve the linked documents, with no time limit. We logged participants' **search interactions** with the *LogUI* framework [217]. Specifically, we logged click events, the ranks and viewpoints of SERP results interacted with, and the time spent on the SERP.

- **Post-Search:** Following the search task, participants could report the identified arguments in a free-text field to simulate the completion of the search task. Further, we asked them to complete the knowledge questionnaire (**knowledge gain**) and report their attitude (**attitude change**) again.

2.3.3. Variables and Analysis

In Table 2.1, we describe the variables used in our study to model relations among searcher and search system, search interaction, and post-search epistemic states. For context on study participants, we captured their age and gender.

To explore the relations between the searcher and search system, search interaction, and post-search epistemic states, we computed the Pearson correlation matrix between all independent and dependent variables that we considered. To explore searcher and search system-dependent group differences through descriptive statistics and ANOVAs, we grouped participants according to their attitude strength (i.e., weak, moderate, strong), the SERP ranking bias condition (i.e., balanced, biased supporting, biased opposing) and categorized them into three levels (i.e., low, moderate, high) of prior knowledge and receptiveness to opposing views, based on the quartiles of the distribution of the respective variable across all participants. We investigate differences between these groups in search interaction (five ANOVAs: attitude-confirming clicks, click rank deviation, number of clicks, time on SERP, hover depth) and in post-search epistemic states (three ANOVAs: directional attitude change, absolute attitude change, knowledge gain). Due to the exploratory and open-ended nature of this study, we did not set a significance threshold but present effect sizes and *p*-values of the ANOVA results as indicators of meaningful relationships that warrant further investigation with confirmatory studies in the future. A table providing an overview of all ANOVA results can be found at the link in Footnote 1.

2.4. Results

We collected data from 280 participants, of which 25 were excluded from the analysis since they failed at least one attention check. Of the 255 remaining participants, 44.3% reported to be female, 54.1% male, and 1.5% preferred not to share their gender. 49% reported to be aged between 18 and 25, 32.5% between 26 and 35, 11% between 36 and 45, 5.9% between 46 and 55, 0.8% between 56 and 65, and 0.8% more than 65 years old.

We aimed for equal distribution of participants across topics and SERP ranking bias conditions: 124 participants were assigned to '*should abortion be legal*' and 131 to '*is obesity a disease*', and 83 to 86 participants were assigned to each of the SERP ranking bias conditions. Amongst the 255 participants, we observed a mean (M) *attitude strength* of 1.8 with a standard error (SE) of 0.06, a mean *prior knowledge* of 0.54 ($SE = 0.01$), and a mean *receptiveness to opposing views* of 0.06 ($SE = 0.02$) with no differences in all three variables between the three SERP ranking bias conditions. Participants, on average, clicked on 2.76 ($SE = 0.16$) search results and spent 7 min 1 sec ($SE = 29.8 sec$) on the search page.

Table 2.2: Subsets of participants considered in the data analysis per dependent variable, since they can only be calculated for participants who fulfilled the requirements. Participants who reported *undecided* attitudes ($n = 19$) were excluded from this analysis due to the inability to determine the direction of *SERP ranking bias* for this group.

Dependent Variables	**Requirements**	**n**
Attitude-Confirming Clicks, Click Rank Deviation	Number of clicks > 0, Attitude Strength > 0	176
Attitude Change, Number of Clicks, Time on SERP, Hover Depth, Knowledge Gain	Attitude Strength > 0	236

Due to varying search interactions, dependent variables were not always applicable across all participants (e.g., confirmation bias was not applicable for participants who did not click on any search result). Thus, we considered different subsets of participants depending on the dependent variable of interest, detailed in Table 2.2. Similarly, the direction of SERP ranking bias (supporting, opposing) cannot be defined for participants who reported to be *undecided* when asked for their attitude ($n = 19$). However, to advance our understanding of the search engine manipulation effect, we explored the attitude change of undecided participants who were exposed to a biased SERP ($n = 13$, see Table 2.5).

Table 2.3: Correlation matrix of all attributes we captured of the searcher and search system, search interaction, and post-search epistemic states, based on participants who had complete data rows ($n = 176$). Positive correlations exceeding 0.1 are colored blue, and negative ones below -0.1 red. Color shades indicate the correlation strength with light shades for weak (0.1 to 0.24) and dark shades for moderate (0.25 to 0.49) correlations. Coefficients are bolded if $p < .05$.

	1.	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**	**9.**	**10.**	**11.**	**12.**
1. SERP ranking bias	1.00											
2. Prior Knowledge	0.02	1.00										
3. Attitude Strength	0.09	0.02	1.00									
4. Receptiveness opp. Views	0.04	-0.09	-0.04	1.00								
5. Attitude-Confirming Clicks	**0.49**	**-0.16**	**0.27**	0.12	1.00							
6. Click Rank Deviation	**-0.37**	0.06	-0.03	0.02	**-0.19**	1.00						
7. Number of Clicks	-0.02	0.14	-0.03	-0.02	-0.14	0.07	1.00					
8. Time on SERP	-0.02	0.03	-0.06	0.02	0.04	-0.04	0.09	1.00				
9. Hover Depth	**-0.30**	0.18	-0.08	-0.02	**-0.22**	**0.39**	**0.26**	**0.10**	1.00			
10. Attitude Change (dir)	-0.06	**0.16**	-0.03	0.01	**0.30**	0.07	0.08	0.00	0.08	1.00		
11. Attitude Change (abs)	-0.02	0.05	**-0.38**	-0.01	**-0.30**	0.02	0.08	0.05	0.02	**-0.41**	1.00	
12. Knowledge Gain	0.02	**-0.26**	-0.07	0.00	0.12	-0.07	0.04	0.07	-0.07	-0.04	0.08	1.00

2.4.1. SEARCH INTERACTION

Attitude-Confirming Clicks We observed weak to moderate correlations between the proportion of attitude-confirming clicks and SERP ranking bias ($r = .49, p < .001$), prior knowledge ($r = -.16, p = .033$), and attitude strength ($r = .27, p < .001$) (Table 2.3). Exploring group differences with an ANOVA, we found variations depending on the SERP ranking bias; $F(2, 107) = 33.74, p < .001, f = 0.79$, with participants exposed to a SERP with an attitude-supporting ranking bias clicking on a higher proportion of attitude-confirming results ($M = 0.72, SE = 0.03$) than participants exposed to a SERP with a balanced ranking ($M = 0.6, SE = 0.02$) or an attitude-opposing ranking bias ($M = 0.28, SE = 0.03$). Further, we observed differences depending on the attitude strength; $F(2,107) = 6.97, p = .001, f = 0.36$, with those who reported having a strong attitude prior to the search clicking on a higher proportion of attitude-confirming results ($M = 0.66, SE = 0.03$) than those who reported having a moderate ($M = 0.54, SE = 0.03$) and weak attitude ($M = 0.42, SE = 0.03$) (see Figure 2.3, A). The ANOVA moreover revealed variations depending on the level of prior knowledge; $F(2,107) = 4.11, p = .019, f = 0.28$, with participants with low prior knowledge clicking on a higher proportion of attitude-confirming search results ($M = 0.66, SE = 0.03$) than participants with moderate ($M = 0.51, SE = 0.03$) and high ($M = 0.46, SE = 0.03$) prior knowledge (see Figure 2.3, B). Lastly, the results indicate an interaction effect between attitude strength and ranking bias; $F(4,107) = 3.27, p = .014, f = 0.35$.

Click Rank Deviation As captured in Table 2.3, we saw a moderate negative correlation between the SERP ranking bias and click rank deviation ($r = -.37, p < .001$). This difference between SERP ranking bias conditions was supported by a between groups ANOVA; $F(2,107) = 15.22, p < .001, f = 0.53$), indicating a higher click rank deviation (i.e., participants clicked on lower-ranked search results) for those exposed to a SERP with an attitude-opposing ranking bias ($M = 0.54, SE = 0.03$) than for those exposed to a balanced ranking ($M = 0.32, SE = 0.02$) or an attitude-supporting ranking bias ($M = 0.27, SE = 0.02$). The ANOVA also revealed a SERP ranking bias and attitude strength interaction; $F(4,107) = 3.26, p = .014, f = 0.35$, where the click rank deviation was lower for participants with strong attitudes compared to those with weak attitudes ($\Delta = -0.24$) if they were exposed to a SERP with an attitude-supporting ranking bias, but higher ($\Delta = +0.14$) if exposed to a SERP with an attitude-opposing ranking bias (see Figure 2.3, C).

Number of Clicks, Time on SERP, Lowest Rank Hovered We saw a moderate negative correlation between SERP ranking bias and hover depth ($r = -.3, p = .008$) (see Table 2.3). Our data did not reveal any correlations between attributes of the searcher and search system and the number of clicks and time spent on the SERP. When exploring group differences designated by attributes of the searcher and search system with ANOVAs, we found that participants' number of clicks varied depending on interactions of the SERP ranking bias with participants' attitude strength; $F(4,163) = 2.73, p = .031, f = 0.26$, as well as their knowledge; $F(4,163) = 3.06, p = .018, f = 0.27$. The lowest ranked result that participants hovered on varied with the SERP ranking bias; $F(4,161) = 4.65, p = .011, f = 0.24$. Further, we saw that this effect of the SERP ranking bias was moderated by participants' attitude strength; $F(4,161) = 2.78, p = .028, f = 0.26$.

2.4.2. Post-Search Epistemic States

Attitude Change We found a weak positive correlation between participants' prior knowledge and their directional attitude change ($r = .16, p = .045$) and a moderate negative correlation between attitude strength and absolute attitude change ($r = -.38, p < .001$) (Table 2.3). We also noticed moderate correlations between participants' attitude-confirming clicks and their directional ($r = .3, p < .001$) and absolute ($r = -.3, p < .001$) attitude change.

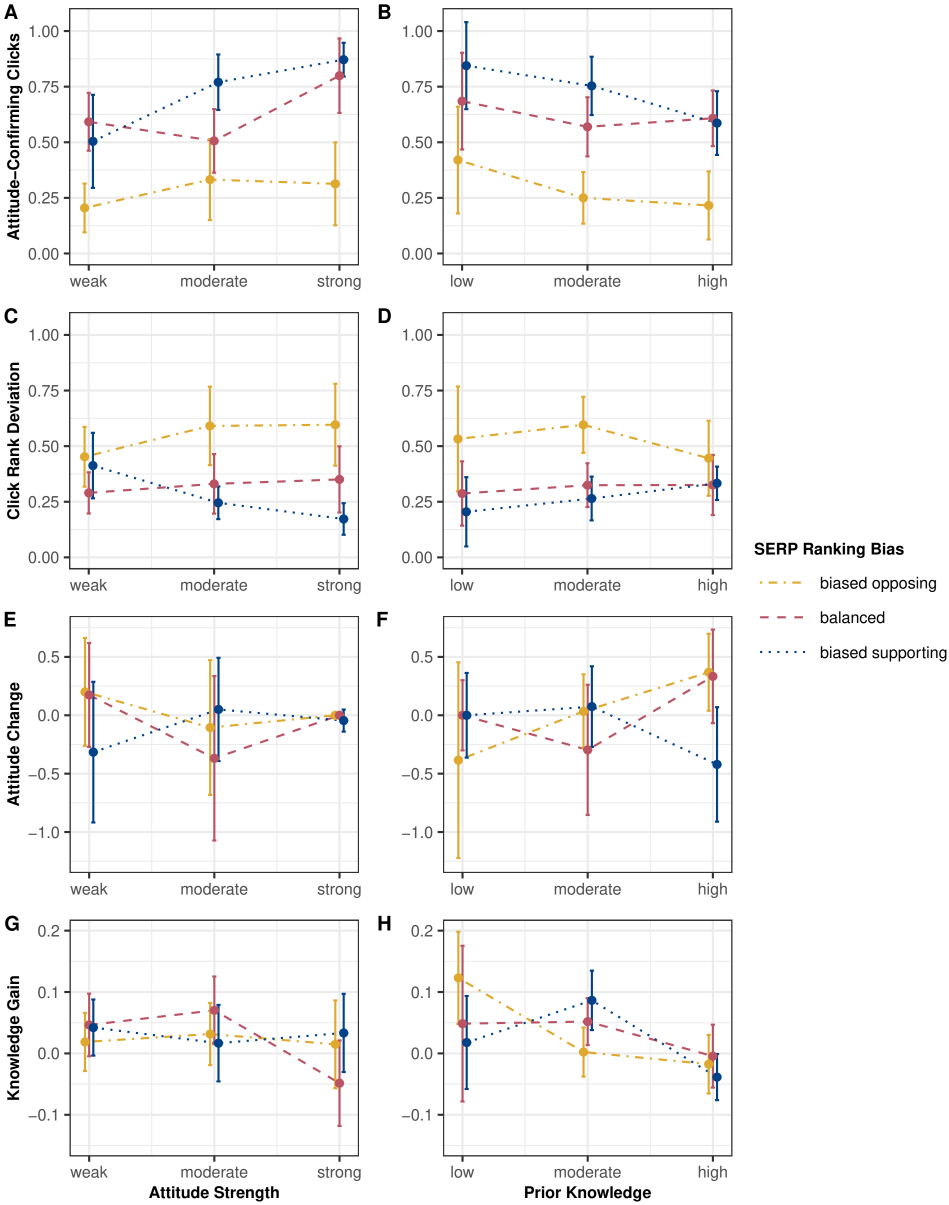


Figure 2.3: **Mean proportion of attitude-confirming clicks** (A, B), **mean click rank deviation** (C, D), **mean attitude change (dir)** (E, F), and **mean knowledge gain** (G, H) per SERP ranking bias condition for different levels of attitude strength and prior knowledge with 95% confidence intervals. Note that we use line plots to illustrate interaction effects.

Table 2.4: **Attitude change of opinionated participants** ($n = 236$)

Attitude Change	**Proportion per SERP ranking bias**		
	supporting	opposing	balanced
changed to opposing	0.12	0.04	0.06
renounced	0.03	0.07	0.04
strengthened	0.17	0.22	0.2
weakened	0.04	0.06	0.05
unchanged	0.64	0.62	0.65

The ANOVA exploring group differences revealed a second-order interaction of SERP ranking bias, participants' knowledge, and their receptiveness to opposing views for directional attitude change; $F(7,163) = 2.55, p = .016, f = 0.33$. When considering *absolute* attitude change, an ANOVA showed group differences depending on participants' attitude strength; $F(2,163) = 16.81, p < .001, f = 0.45$: those with weak ($M = 0.79, SE = 0.05$) and moderate attitudes ($M = 0.65, SE = 0.07$) were more likely to change their attitudes, whereas those with strong attitudes were highly unlikely to change their attitudes ($M = 0.08, SE = 0.03$) (Figure 2.3, E). For absolute attitude change, an interaction of the SERP ranking bias and participants' knowledge was found; $F(4,163) = 2.75, p = .03, f = 0.26$ (Figure 2.3, F).

We investigated the proportions of participants who adopted the majority viewpoint amongst the high-ranked SERP results. We did not find evidence for the search engine manipulation effect since the proportion of participants who changed their attitude towards the opposing attitude was lower for those exposed to a search page with attitude-opposing (4%) than for attitude-supporting ranking bias (12%) and balanced ranking (6%) (see Table 2.4). Additionally, we explored attitude change for participants who reported being undecided before the search task and were exposed to a viewpoint-biased SERP ($n = 13$). In this group, the same proportion of participants (31%) adopted the majority viewpoint as the minority viewpoint (see Table 2.5).

Knowledge Gain We found a moderate negative correlation between participants' level of prior knowledge and their knowledge gain ($r = -0.26, p < .001$). An ANOVA alludes to differences between participants with varied levels of both prior knowledge; $F(2,163) = 9.69, p < .001, f = 0.34$, and attitude strength; $F(2,163) = 3.28, p = .04, f = 0.20$. Participants with low prior knowledge gained knowledge ($M = 0.07, SE = 0.009$), those with high prior knowledge did not ($M = -0.02, SE = 0.005$) (see Figure 2.3, H). Further, participants with strong prior attitudes were less likely to gain knowledge ($M = 0.0, SE = 0.008$) than those with moderate ($M = 0.04, SE = 0.008$) and weak attitudes ($M = 0.03, SE = 0.008$) (see Figure 2.3, G).

Table 2.5: **Attitude change of undecided participants in biased SERP ranking conditions** ($n = 13$)

Attitude Change	Proportion
still undecided	0.38
adopted majority viewpoint	0.31
adopted minority viewpoint	0.31

2.5. Discussion

Reflecting on this exploratory user study, we discuss and contextualize key findings and their implications, as well as acknowledge limitations and outline avenues for future work.

Confirmation and Position Bias. Our data revealed that confirmation bias was strongest in participants with low knowledge or strong attitudes. Both confirmation and position bias were affected by the SERP ranking, where participants exposed to an attitude-supporting ranking bias displayed higher confirmation and position bias than those exposed to an attitude-opposing ranking bias. Participants in the attitude-opposing ranking bias condition tended to click on lower-ranked results, and the proportion of clicks on attitude-confirming results, while diminished, was still substantial. This suggests that participants deliberately sought not only results that were not attitude-opposing (i.e., neutral results) but more so, attitude-confirming results within the lowest ranks of the SERP. In contrast, those exposed to attitude-supporting ranking bias engaged less with low-ranked results. From this, we infer that the desire to find information confirming prior attitudes (confirmation bias) rather than to explore diverse perspectives caused this behavior in participants in the attitude-opposing ranking bias condition. This aligns with previous findings on search interactions being not only shaped by algorithms, determining exposure through selection and ranking but also by users, tailoring interaction to maintain prior beliefs [302, 323]. Yet, Robertson et al. [302] state that this does not imply that exposure biases are less concerning than previously suggested but that they might cause more indirect effects persistently over time. Furthermore, our results point towards a potential risk arising from attitude-supporting as opposed to attitude-opposing exposure bias; while attitude-supporting ranking bias does not manipulate individuals into changing their attitudes (i.e., the search engine manipulation effect), it amplifies their interaction biases, thus hindering responsible opinion formation.

Answering **RQ1**, we found that searcher and system attributes significantly shape search interaction, with confirmation bias being strongest in participants with low knowledge or strong attitudes, as well as those exposed to an attitude-confirming ranking bias, and position bias being weakest for participants who were exposed to an attitude-opposing ranking bias. Although exposure bias considerably influences search behavior, confirmation bias drives user interaction to somewhat diminish the impact of attitude-opposing, but not attitude-supporting, exposure bias.

Attitude Change & Knowledge Gain. A salient trend regarding attitude change showed that participants with strong prior attitudes were highly unlikely to change their attitude, regardless of viewpoint exposure biases or prior knowledge levels. SERP ranking bias, on the other hand, does not seem to affect attitude change. When exploring attitude change in a more nuanced manner, by comparing the prevalence of different attitude change categories across the SERP ranking bias conditions, we did not find evidence of attitude change that would indicate an effect of exposure bias like the search engine manipulation effect. Across SERP ranking bias groups, most participants maintained their initial attitude, not aligned with the observations reported by Epstein and Robertson [82]. However, in this study, the authors considered attitude change for voting decisions, by investigating the impact of search results that favored one candidate in an election over a different candidate. Strong attitudes on debated topics are often rooted in stable moral values [140] and might be less prone to change than election decisions that were found to be impacted by less stable candidate qualities [45]. Drawing from our findings, we infer that the impact of exposure biases on searchers' attitudes towards debated topics, as well as search interactions, is likely mediated by nuanced pre-search epistemic states related to users' attitudes, such as their *importance*, *moral conviction*, or *certainty* [140, 172, 345]. Interestingly, our data indicated that participants with strong attitudes were less likely to gain knowledge than those with moderate and weak attitudes. Corroborating similar findings from prior work, participants with relatively high prior knowledge were less likely to gain more knowledge [93, 306].

Addressing **RQ2**, we discovered that searchers' attitude strength and prior knowledge, but not exposure effects, impact attitude change and knowledge gain. Searchers with relatively strong attitudes were less inclined to change their attitudes and were less likely to gain new knowledge, as did searchers with relatively high prior knowledge.

Regarding **RQ3**, our results showed that attitude change is linked to the level of confirmation bias, with position bias or search effort having no significant impact. Specifically, individuals who exhibited higher confirmation bias in their search interactions were less likely to change their attitudes compared to those with lower confirmation bias. Similar to observations by Draws et al. [74], these findings suggest that the influence of search results on attitude change is primarily driven by selective user interaction rather than mere exposure. These observations expand on previous findings on the searcher-system interplay of exposure and interaction, revealing not only its impact on search behavior but also on attitude change. Our study did not uncover any evidence indicating that search interaction influenced knowledge gain.

Additional Observations. We noted first- and second-order interaction effects of attributes of the searcher and search system on search interaction and post-search epistemic states. We found interaction effects of SERP ranking bias and attitude strength on attitude-confirming clicks, click rank deviation, number of clicks, and hover depth, and interaction effects of SERP ranking bias and prior knowledge on the number of clicks and attitude change. Although our sample size was not sufficient to capture further higher-order interactions, we interpret these effects as indicators that exposure effects on search interaction as well as post-search epistemic states are likely shaped by prior knowledge and attitude. This underlines the need to investigate the role of more nuanced pre-search epistemic states related to searchers' attitudes in web search on debated topics, such as their *importance*, *moral conviction*, or *certainty* [140, 172, 345].

Apart from a second-order interaction effect on directional attitude change, we did not see any relations of searchers' *receptiveness to opposing views* to their search interaction and post-search epistemic states. This might be due to the focus of this user trait on engagement with and interpretation of information in the context of *passive exposure* [229]. However, individuals tend to turn to web search when *actively seeking* for information on a topic (locate, select, and access sources) [314]. Given the important role of active user interaction choices as opposed to passive exposure, a user trait that captures not only their receptiveness but also whether they tend to actively seek out information with opposing views likely plays a more prominent role in search on debated topics (e.g., *intellectual humility* [104, 267, 294]).

2.5.1. IMPLICATIONS

In our study, we observed that viewpoint biases in exposure and interaction shape web search on debated topics and impede searchers from closing their knowledge gaps on the topic. Our data did not indicate a direct relation between exposure bias and attitude change. In particular, we did not observe a shift towards the opposing attitude in conditions of attitude-opposing ranking bias, i.e. the search engine manipulation effect, as was previously reported [82]. However, when aiming for search interactions that enable responsible opinion formation, our findings show that we should place equal or even greater emphasis on the effects of exposure bias aligned with user attitudes. Participants who were exposed to attitude-supporting ranking bias exhibited particularly high confirmation and position bias and, in turn, a low likelihood of engaging with attitude-opposing results, behavior that impedes well-informedness and responsible opinion formation. Yet, from the perspective of search engines that optimize for user satisfaction, their behavior of engaging primarily with highly-ranked results could be interpreted as a signal for well-tailored relevance criteria [1, 149]. This underscores that, when dealing with debated topics, personalization by optimizing relevance criteria with the objective of increasing satisfaction of the individual user can reinforce their biases [22, 44], and thus hinder engagement with diverse viewpoints [105, 296].

Recently, there have been calls for improved search systems that provide better support for complex information needs [320, 325]. During search on debated topics, searchers could be better supported in closing their knowledge gaps and engaging with diverse viewpoints [105, 296]. Given the role of user interaction noted in this and other studies [74, 302, 323], approaches that aim to provide better support for complex information needs should not only facilitate access to diverse viewpoints but prompt and empower searchers to productively interact with results that advance their level of informedness. Our findings of knowledge- and attitude-related differences suggest that for such interventions to effectively support productive searches on debated topics, they would likely need to be personalized. This raises the question on *which searchers* would benefit from *what* kind of support and *when*.

Individuals with strong attitudes and limited knowledge, who, according to our observations, exhibited the most pronounced confirmation bias, might need interventions that motivate engagement with attitude-opposing results (e.g., warning labels [297]) but prove to be challenging to reach, regardless of the intervention. Individuals with high prior knowledge, on the other hand, exhibited low interaction bias. Yet, they might have interacted with attitude-opposing results to counter-argue and discount their content [60, 335], instead of objectively assessing it to close knowledge gaps. These searchers might not require support to interact with attitude-opposing information but instead to objectively assess their content (e.g., boosting strategies for informed search [32] or intellectual humility [294, 295]). While these findings of knowledge- and attitude-related differences suggest the need for personalized interventions to support search for responsible opinion formation, such interventions would require data on an individual's views, raising privacy concerns. Consequently, this would have to be approached in a privacy-aware manner, e.g., by ensuring that users can understand and control what shapes their user model and how it affects the information environment [344].

While this study focuses on web search, understanding the effects of exposure, interaction, and their interplay is highly relevant to similar domains of web interactions with information that can impact individuals' opinions, such as news recommenders, social media platforms, or discussion forums. Recent research on how individuals from Gen Z (born between 1997 and 2012) engage with information online revealed that they tend to encounter rather than actively search for information and that their interactions are strongly driven by social motivations rather than by truth-seeking [125]. However, Hassoun et al. [125] remarked that Gen Z highly values *information sensibility* which the authors define as *a socially-informed awareness of the value of information encountered online*. This could serve as a motivating factor for cultivating web interactions for responsible opinion formation.

2.5.2. Caveats, Limitations, and Future Work

As with all empirical and exploratory research, our study is not without limitations. We framed the information search task in a single search session with preselected, viewpoint-annotated SERP results and focused on participants' clicking behavior. Factors such as querying, query refinement, or multiple search sessions were beyond the scope of this work. Nonetheless, the qualitative feedback we collected was mostly positive and did not indicate any frustration regarding the restricted interaction options. Although we considered two topics to represent different levels of controversy, we plan to consider diverse tasks and additional topics to enhance the scope of generalizability of our findings. Future research should also gauge the long-term impact of the search session on attitude, knowledge, and decision-making by conducting confirmatory follow-up studies, enabling a more realistic search process, expanding the range of topics and tasks examined, and investigating the mediating effects of search interaction measures.

Cognitive biases arising from the task design of crowdsourced user studies can negatively impact data quality [80]. We assessed potential biases in our study with the Cognitive Biases Checklist by Draws et al. [71], identifying that various cognitive biases could have affected the search interaction data by causing participants to diverge from their usual behavior in a real search setting. The pre-search knowledge questionnaire might have caused *anchoring effects* by leading participants to search for answers to the knowledge questions they encountered. Further, the time invested in filling the pre-search questionnaires could have resulted in a *sunk cost fallacy* of sticking to and rushing through the task to receive the final reward, even though participants may not be genuinely interested in the search task. To counter this potential bias, we encouraged genuine behavior within the search task with an added incentive of a bonus payment if they successfully identified three high-quality arguments in the search session.

We used knowledge questionnaires to capture the general depth of knowledge among searchers. However, we came to realize that one-dimensional knowledge questionnaires that capture the depth of topic knowledge do not suffice to measure *informedness* in the context of debated topics. We plan to broaden the scope of our measures to include the breadth of knowledge, i.e., encompassing knowledge about different viewpoints.

2.6. CONCLUSION

We presented the results of our user study to explore the relations between attributes of the searcher and search system, search interaction, and user post-search epistemic states—a first step towards developing a comprehensive and user-focused understanding of web search on debated topics. Our insights can inform the design of interventions that support responsible opinion formation. We observed that search interaction was shaped by search system exposure, attitude strength, and prior knowledge. Attitude change was not directly influenced by search system exposure but instead by participants' confirmation bias and their initial attitude strength. Our findings suggest that the effects of exposure and interaction biases, as well as their interplay on post-search epistemic states, likely depend on nuanced epistemic states related to searchers' attitudes. The knowledge and attitude-dependent differences suggest that interventions to support fruitful search interactions likely require privacy-conscious personalization, adapting the intervention to users' pre-search epistemic states. These insights further underscore that customizing search rankings based on implicit feedback to enhance user satisfaction can have harmful repercussions in the context of debated topics, as it is prone to cultivate attitude-supporting exposure bias, thereby reinforcing confirmation bias and hindering responsible opinion formation. Our findings could extend to other web interactions involving exposure to algorithmically curated information that has the potential to influence opinions, such as interactions on social media platforms or with news recommender systems. This is particularly of interest in the era of transition from the traditional linear information journey—from queries to answers—to more fluid and socially-oriented journeys that alternate between information-encountering and information-seeking [125].

3

Unveiling Perspectives: Leveraging Stance Labels for Increased Ranking Transparency

The user study described in this chapter was planned and conducted in collaboration between Alisa Rieger, Tim Draws, Zhangyi Wu, Federico Cau, with input from Francesco Barile and Nava Tintarev. Specifically, Alisa and Tim guided the research design and collaborated with Zhangyi in planning the material preparation, user study methodology, and implementation. The material preparation of the dataset of search results with stance labels and explanations was primarily done by Zhangyi with Tim's guidance. In a joint effort, Zhangyi, Alisa, Tim, and Federico implemented the user study. The data collection was carried out by Zhangyi and Alisa. Parts of this user study have been disseminated in [382]. Alisa conducted the data analysis presented in this chapter, which she wrote with the guidance of Maria Soledad Pera.

3

3.1. Introduction

The opacity of web search engines can impede searchers from detecting potential bias or incompleteness in the presented information. This, in turn, could prevent them from engaging with diverse viewpoints to achieve well-rounded knowledge gain, which is essential for responsible opinion formation [226].

To increase the transparency of search rankings and advance search systems to prioritize human learning and information literacy, Smith and Rieh [325] suggest enriching the knowledge context. This can be achieved by providing meta-information, for instance, *epistemic cues* that indicate the epistemic qualities of information [210]. Such cues can provide information on the context or content level, e.g., which sources cite a given article, or whether an article provides primarily facts or opinions. For search on debated topics, Draws et al. [70] suggested displaying content-related epistemic cues by providing automatically generated labels that visualize the stance of each search result (i.e., neutral, in favor, against) with explanations to achieve transparency of viewpoint ranking biases and facilitate and motivate engagement with diverse viewpoints.

Draws et al. [70] have laid the groundwork for this intervention, observing a satisfying quality of explainable stance detection methods for search results. Nevertheless, the exploration of its effects on users' search behavior remains a pending endeavor, leading us to pose the following research question:

RQ1: Do automatically generated stance labels and explanations of the labels for search results increase the diversity of viewpoints users engage with, even if search results are biased towards one particular viewpoint?

We addressed this research question with a preregistered between-subjects user study (N=198) simulating an open search task on debated topics.[1] In this task, participants were asked to search for information on a selected debated topic. We manipulated (i) whether the search result viewpoint ranking was neutral or biased, and (ii) whether search results were displayed regularly (without stance labels), with stance labels, or with stance labels and explanations. To investigate RQ1, we logged participants' search interactions and calculated the viewpoint diversity of their clicks.

Our results indicate that stance labels increase the viewpoint diversity in users' engagement and mitigate the effects of ranking bias that we observed for regular search result pages without stance labels. We did not find differences between the effects of merely stance labels compared to stance labels with explanations. From our findings, we conclude that providing more transparency over the viewpoint ranking by displaying stance labels is a valuable strategy to facilitate and motivate fruitful web search on debate topics.

3.2. Related Work

Search engine result pages (SERPs) generated in response to search queries pertaining to debated topics are often viewpoint-biased, overrepresenting one viewpoint over

[1] The pre-registration is openly available at https://osf.io/3nxak, and the user study data at https://doi.org/10.5281/zenodo.10993022.

others in the highly-ranked search results [72]. This can, directly and indirectly, impact not only users' interactions (e.g., preventing engagement with diverse viewpoints) but even their attitudes on the topics [74, 82, 302]. For instance, searchers can experience the *search engine manipulation effect* when changing their attitude towards the overrepresented viewpoint [82, 83]. In the context of debated topics, search interactions can be further impacted by cognitive biases, such as the confirmation bias, i.e., individuals who tend to engage with search results that confirm their attitude [18, 244]. It is important to consider, however, that users are frequently unaware of such biases [18, 99]. The lack of transparency in search rankings obstructs their understanding of the knowledge space, identification of viewpoint-biased SERPs, and evaluation of their interactions, for instance in terms of their engagement with diverse viewpoints [226, 320].

Previous research has proposed and explored various approaches to support users in navigating search on debated topics and overcoming these challenges. Researchers have suggested approaches to re-rank search results to reduce viewpoint bias on SERPs [72]. While this constitutes an important step towards supporting engagement with diverse viewpoints, the process of re-ranking necessitates normative assessments about the notion of viewpoint diversity according to which the results should be re-ranked [129] and requires decisions on value trade-offs (i.e., serving user needs vs. democratic values) [48]. Other research has investigated *argument retrieval systems* that support searchers in gaining knowledge of various arguments by retrieving and directly presenting different arguments from search results [3, 38, 366]. Yet, directly providing information instead of pointers of where to look for it runs the risk of further diminishing transparency and could reinforce searchers' inclination to believe that they do not need to invest cognitive effort, rather than empower them to thoroughly explore the information space and differentiate among sources [320, 325].

Recently, Draws et al. [70] proposed a strategy to increase the transparency of search rankings for queries on debated topics by displaying stance labels for search results. This strategy would sidestep the risks and limitations associated with the aforementioned approaches by providing cues for more informed user choices while eliminating the need for decisions on value trade-offs. The authors found that stance labels for search results can be generated with explainable, automatic, target stance detection methods and conducted extensive quality evaluations, including a user study to determine the meaningfulness of stance labels and explanations for users. Draws et al. [70] found that transformer-based models such as *DistilBERT* yielded high predictive quality and explanation methods such as *LIME* with salience-based visualizations generated compelling explanations to users. However, the effect of stance labels with explanations on search behavior has yet to be explored.

3.3. METHOD

To advance understanding of whether search behavior is influenced by the transparency of the viewpoint ranking, using stance labels without and with explanations, we conducted a randomized controlled trial between-subjects design, comparing **clicking diversity** (i.e., the diversity of viewpoints clicked on) between

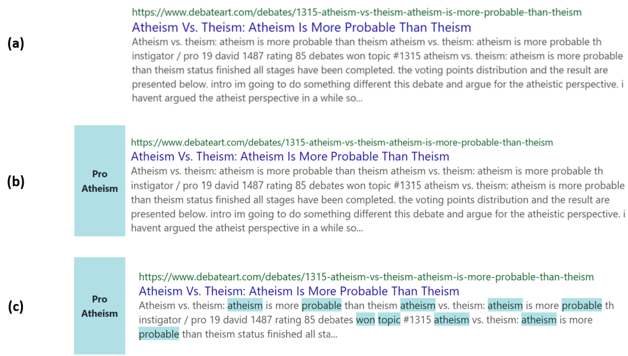


Figure 3.1: **Search result display.** Example search result displayed in (a) regular, (b) stance label, and (c) stance label and explanation search result display condition.

six groups. We manipulated the factors **SERP ranking bias** *(biased towards the attitude-opposing viewpoint, neutral)* and **SERP display** *(regular, with stance labels, with stance labels and explanations)* and tested the following hypotheses:

- **H1:** Users who are exposed to a SERP with a viewpoint-biased ranking interact with less diverse results than users who are exposed to a SERP with a neutral ranking.
- **H2:** Users who are exposed to a search interface with (1) stance labels or (2) stance labels with explanations for each search result interact with more diverse search results than users who are exposed to a regular search interface without stance labels.
- **H3:** Users who are exposed to a search interface with (1) stance labels or (2) stance labels with explanations are less susceptible to the effect of a viewpoint-biased ranking on clicking diversity.

3.3.1. Experimental Setup

Dataset. For the three search result display conditions (regular, stance label, stance label and explanation), we required a dataset with search results on debated topics, stance labels (*against*, *neutral*, and *in favor*), and stance label explanations. To the best of our knowledge, datasets containing search results with stance labels and explanations are not publicly available. Hence, we built our own, starting from a public dataset by Draws et al. [72] which contains URLs, titles, snippets, and stance labels for 1475 search results on three debated topics that affect either the user (*atheism*), businesses (*intellectual property rights*), or society (*school uniforms*) and offer valid arguments for both supporting and opposing viewpoints.

To create stance labels and corresponding explanations, we adopted the approach used in [70]: we mapped the stance labels from the provided dataset—obtained through human expert annotation using a seven-point Likert scale, indicating varying degrees of opposition and support—into the three categories: *against*, *neutral*, and *in favor*. Then, we built a cross-topic model, fine-tuning and evaluating the uncased DistilBERT model [311] on the search result corpus, which we were able to retrieve for 1125 search results. For pre-processing, we combined each search result's title, snippet, and head and tail sections, making sure the final text was 510 tokens long. By applying a stratified split to maintain the distribution of topics and labels across the subsets, we allocated 75% of the search results for training and 25% for validation and testing, achieving a final macro F1 score of 0.72 on the test data. For the explanations corresponding to the stance labels, we proceeded with 108 search results from our test dataset, for which stance was predicted correctly. We obtained feature attributions for the predictions by applying LIME [292] and created salience-based explanations over text (i.e., using colors to highlight features based on feature importance), showing the 20 most important tokens.[2]

Search Interface. We designed the search interface to embody conventional search interfaces and presented search results, including title, url, and snippet, in ordered lists with ten search results on a page. We randomly selected 21 search results per topic from our dataset (seven against, seven neutral, and seven in favor). Depending on the assigned ranking bias condition, participants either saw a *biased* or *neutral* SERP, displaying randomly selected search results from the different stance groups to fulfill fixed ranking templates (see Table 3.1). On the biased SERP, a majority of search results with a stance opposing the participant's attitude were presented, whereas the neutral SERP presented a majority of search results with a neutral stance. In the event that participants issued a new query, the order of search results was shuffled in a semi-random manner, following the predefined ranking templates.

Depending on the assigned search result display condition, participants would see the search results either (a) regularly, (b) with stance labels, or (c) with stance labels and explanations (see Figure 3.1).

3.3.2. VARIABLES

Independent variables.

- *SERP ranking bias.* Biased (majority of attitude-opposing search results), or neutral (majority of neutral search results) (see Table 3.1).

- *SERP display.* Regular, with stance labels, or with stance labels and explanations (see Figure 3.1).

Dependent variable.

- *Clicking Diversity* The viewpoint diversity of participants' clicks was calculated with the Shannon Index [321]. The result is a single value that represents the evenness in distribution over different categories (clicks on neutral,

[2]For comprehensive details on the dataset creation process, please refer to the preregistration at https://osf.io/3nxak.

Table 3.1: Templates for the first SERP (i.e., the top 10-ranked search results) in each SERP ranking bias condition.

Rank	Biased	Neutral
1	Attitude-Opposing	Neutral
2	Attitude-Opposing	Neutral
3	Attitude-Opposing	Neutral
4	Attitude-Opposing	Neutral
5	Attitude-Confirming	Attitude-Opposing
6	Neutral	Neutral
7	Attitude-Opposing	Attitude-Confirming
8	Attitude-Confirming	Attitude-Opposing
9	Neutral	Neutral
10	Attitude-Opposing	Attitude-Confirming

3

attitude-opposing, and attitude-confirming search results). Higher values indicate more uniform distributions and, consequently, higher viewpoint diversity of participants' clicks.

We calculated clicking diversity with the Shannon Index as $-\sum_{i=0}^{2} \frac{p_i}{N} \ln(\frac{p_i}{N})$, with N being the total number of clicks made in one session, and p_0, p_1, p_2 the number of clicks of *neutral*, *attitude-opposing*, and *attitude-confirming* search results, respectively. For instance, clicking diversity for 0 *neutral*, 1 *attitude-opposing*, and 3 *attitude-confirming* clicks is $0 + 0.34 + 0.21 = 0.55$. Given that we consider three categories, the values range from 0 (only one category clicked) to 1.1 (the same number of clicks in all three categories).

Exploratory variables.

- *Clicks on Neutral / Attitude-Opposing / Attitude Confirming Items.* The proportion of a participant's clicks on search results with a neutral / attitude-opposing / attitude-confirming stance.

- *Mean Click Viewpoint.* To calculate the mean viewpoint clicked on, we considered clicks on search results with an attitude-opposing stance as -1, clicks on search results with an attitude-confirming stance as +1, and clicks on neutral search results as 0. For instance, the mean click viewpoint for 0 *neutral*, 1 *attitude-opposing*, and 3 *attitude-confirming* clicks is $(0 - 1 + 3)/3 = 0.66$.

- *Number of Clicks.* Number of search results participants clicked on.

- *Task Completion Time.* Time participants spent on the search task.

- *Mean Click Rank.* Mean rank of search results clicked on. E.g., if a participant clicked on search results displayed on the first, the fourth, and the sixth rank, the mean click rank would be $(1 + 5 + 6)/3 = 4$.

- *Attitude Change*. Difference between attitude reported in the pre- and the post-search questionnaire. Negative values indicate that participants reported a weakened attitude, or changed their attitude to the opposing, while positive values indicate that participants reported a strengthened attitude. Since we only recruited participants with strong prior attitudes, the values for attitude change could range between -6 (*strongly supporting* to *strongly opposing* or vice versa) and +1 (*supporting* to *strongly supporting* or *opposing* to *strongly opposing*)

- *Topic*. One of atheism, intellectual property rights, or school uniforms.

3

We additionally captured variables such as participants' gender and age to describe the representativeness of the sample.

3.3.3. PROCEDURE

We collected the questionnaire data via the online survey platform *Qualtrics* [283], from which we directed participants to our search interface for the search task. We recruited participants via *Prolific* [277]. Participants had to be at least 18 years old and be fluent English speakers. The following procedure for data collection was approved by the ethics committee of our institution:

Pre-search. After giving informed consent to participate in the study, we asked participants to report their demographics. We then asked them to state their attitudes on the three debated topics on a seven-point Likert scale. We assigned participants to one of the three debated topics (i.e., *atheism*, *intellectual property rights*, and *school uniforms*) for which they reported having a strong attitude (i.e., strongly opposing, opposing, supporting, or strongly supporting). Participants who did not report a strong attitude on either of the topics could not participate in the search task and were therefore paid proportional to the time they invested. If a participant reported strong attitudes on multiple topics, they were assigned to the topic that had the fewest participants at that point in time (i.e., to achieve a balanced topic distribution). We semi-randomly assigned them to one of the two SERP ranking bias conditions and one of the three SERP display conditions, aiming for equal numbers of participants among the six groups.

Search. We asked participants to imagine the following scenario: *You and your friend were having dinner together. Your friend is very passionate about TOPIC and couldn't help sharing his views and ideas with you. After the dinner, you decide to further inform yourself on the topic by conducting a web search.*

After reading the scenario, participants could proceed to the search interface. Here, they could enter queries into a search bar. When the query contained topic-related phrases they were exposed to search results randomly selected from the set of 21 available search results on the assigned topic (i.e., 7 against, 7 neutral, and 7 in favor), ranked according to the assigned ranking condition template (see Figure 3.1). Depending on the SERP display condition participants were assigned to, they were presented either with regular search results, search results with stance labels, or search results with stance labels and explanations (see Figure 3.1). They could explore the search results by clicking on them to retrieve the linked web pages.

Table 3.2: Distribution of participants across conditions of search result display (lines) and SERP bias (columns).

	Biased Ranking	Neutral Ranking
Regular	36	35
Stance Label	30	33
Stance Label and Explanation	34	30

Post-search. After completing the search, we asked participants to report their post-search attitude toward the assigned topic. Concluding the study, we debriefed participants about the purpose of the study.

3.3.4. DATA ANALYSIS

Before data analysis, we excluded data generated by participants who did not meet the preregistered inclusion criteria by failing the attention checks, not accessing the search platform, or not clicking on any links during the search. To test the hypotheses, we conducted an ANOVA, investigating the main and interaction effects of the two independent variables (i) *SERP display* (regular, with stance labels, with stance labels and explanation) and (ii) *SERP ranking bias* (biased, neutral) on *clicking diversity* (H1, H2, H3). Aiming at a type 1 error probability of $\alpha = 0.05$ and applying Bonferroni correction to correct for multiple testing, we set the significance threshold to $\frac{0.05}{3} = 0.017$. In addition to the analyses described above, we conducted additional exploratory analyses to advance the broader understanding of engagement with different viewpoints, search effort, and attitude change during search on debated topics.

3.4. RESULTS

Prior to data collection, we computed a required sample size of 205 participants based on an effect size of $f = 0.25$, a significance threshold $\alpha = \frac{0.05}{3} = 0.017$ (due to testing three hypotheses), a desired power of $(1-\beta) = 0.8$ and given that we tested, depending on the hypothesis, six groups (i.e., three SERP display conditions and two SERP ranking bias conditions) for a between-subjects ANOVA. 198 participants passed the inclusion criteria. They were roughly equally distributed across the three topics (65-67 participants per topic) and conditions (see Table 3.2). Participants, on average, clicked on 3.4 ($SE = 0.16$) search results and spent 6 min 6 sec ($SE = 15.5 sec$) on the search page. Completion of the task was rewarded with £1.10 (mean across all participants = £12.80/h), fulfilling the Prolific recommendations for ethical rewards.

Engagement with Different Viewpoints. To test H1, H2, and H3, we conducted an ANOVA, comparing clicking diversity between the two SERP ranking bias and three SERP display conditions. Given the Bonferroni corrected significance threshold of 0.017, we did not find evidence for an effect of the SERP ranking on clicking diversity ($F(1,192) = 5.46, p = .02, f = 0.17$), and thus to support H1. However, confirming H2, we did find evidence for an effect of the SERP display on clicking diversity ($F(2,192) = 8.18, p < .001, f = 0.29$), whereas participants exposed to SERPs

Table 3.3: **Engagement with Different Viewpoints.** Means and standard errors of clicking diversity and mean click viewpoint per SERP ranking and SERP display condition.

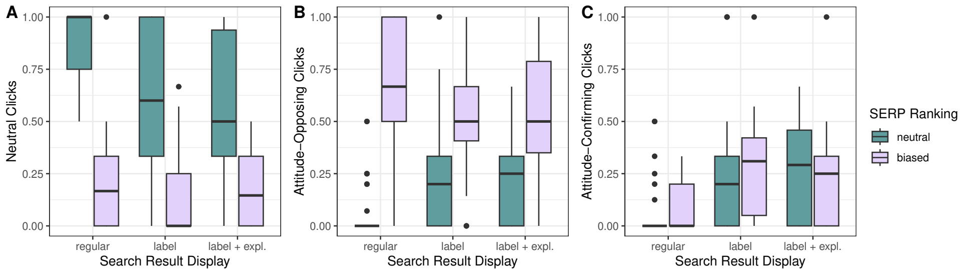
SERP Ranking	SERP Display	Clicking Diversity *mean*	Clicking Diversity *SE*	Mean Click Viewpoint *mean*	Mean Click Viewpoint *SE*
neutral	reg	0.22	0.05	-0.009	0.03
	lab	0.54	0.08	0.005	0.06
	lab + expl	0.64	0.08	0.02	0.05
biased	reg	0.52	0.07	-0.61	0.05
	lab	0.62	0.07	-0.28	0.08
	lab + expl	0.64	0.07	-0.34	0.08

3

with warning labels, as well as warning labels and explanations, clicked on more diverse search results than participants who were exposed to regular SERPs (see Table 3.3). We did not find evidence for an interaction effect between SERP display and SERP ranking on clicking diversity ($F(2,192) = 2.37, p = .1, f = 0.16$), and thus to support H3.

To gain further insights into the viewpoints participants engaged with, we additionally explored differences between clicks on neutral, attitude-opposing, and attitude-confirming search results (see Figure 3.2) and the mean click viewpoint between the six groups (see Table 3.3). Our explorations indicated that the mean click viewpoint was affected by both SERP ranking ($F(1,192) = 65.92, p < .001, f = 0.59$) and SERP display ($F(2,192) = 4.65, p = .011, f = 0.22$), as well as their interaction ($F(2,192) = 3.42, p = .035, f = 0.19$). Participants exposed to a biased ranking clicked on a higher proportion of attitude-opposing search results than those exposed to a neutral ranking. Further, from the participants in the biased ranking condition, those exposed to regular search results clicked on a higher proportion of attitude-opposing search results than those exposed to search results with stance labels without and with explanation. Exploring the distribution of clicks on search results with different viewpoints across SERP display conditions depicted in Figure 3.2, we observed that independent of the SERP ranking condition, participants primarily engaged with the over-represented viewpoint (i.e., neutral in neutral ranking and attitude-opposing in biased ranking). The data further indicates, that both stance labels and stance labels with explanations equally reduce engagement with the over-represented viewpoint while increasing engagement with the under-represented viewpoints. We did not observe differences in engagement with different viewpoints between the three topics.

Search Effort. We aimed to explore whether the interventions intended to provide increased viewpoint ranking transparency to facilitate thorough search behavior, affected the effort participants invested in the search task. For that, we explored differences between the SERP display conditions in the number of clicks, the task completion time, and the mean click rank (see Table 3.4). Exploring the mean values, we observe reduced mean clicks and elevated mean click ranks for participants exposed to SERPs with stance labels and stance labels and explanations. We also noted prolonged mean task completion times for those exposed to SERPs



3

Figure 3.2: **Clicks on search results with different viewpoints.** Boxplots displaying the distribution of participants' click proportion on search results with an (A) neutral, (B) attitude-opposing, (C) attitude-confirming stance per search result display (regular, with stance label, with stance label and explanation) and ranking (neutral, biased) condition.

Table 3.4: **Search Effort.** Means and standard errors of the number of clicks, task completion time, and mean click rank per SERP ranking and SERP display condition.

		Number of Clicks		**Task Completion Time**		**Mean Click Rank**	
SERP Ranking	SERP Display	*mean*	*SE*	*mean*	*SE*	*mean*	*SE*
neutral	reg	3.29	0.45	338	27	3.09	0.27
	lab	2.94	0.32	348	23	4.48	0.34
	lab + expl	2.93	0.3	391	69	4.35	0.36
biased	reg	3.97	0.3	338	32	4.17	0.3
	lab	3.57	0.32	371	33	4.76	0.45
	lab + expl	3.53	0.55	418	36	4.57	0.52

with stance labels and explanations. Our explorations did not reveal any differences in search effort between topics.

Attitude Change. Our explorations of participants' attitude change did not indicate any differences between the SERP ranking and SERP display conditions. Thus, we did not find evidence for the search engine manipulation effect. Further, we did not observe correlations of attitude change with participants' search behavior and engagement with different viewpoints. However, the exploratory insights indicate a difference between topics ($F(2,195) = 3.28, p = .04, f = 0.18$). Specifically, participants engaging with search results on *atheism* were less likely to change their attitude ($Mean = -0.96, SE = 0.13$) than those engaging with search results on *school uniforms* ($Mean = -1.39, SE = 0.13$).

3.5. Discussion

We conducted a user study to test whether increased viewpoint ranking transparency affects search behavior during web search on debated topics. For that, we investigated the effect of stance labels without and with explanations, indicating the stance of each search result in conditions of viewpoint-biased and neutral SERP rankings on the diversity of viewpoints participants engaged.

Findings on Engagement with Different Viewpoints. We found evidence that stance labels and stance labels with explanations lead to increased clicking diversity, confirming H2. We did not find evidence for an effect of SERP ranking (H1), and an interaction effect with SERP display (H3) on clicking diversity.

Our explorations of the distinct viewpoints participants clicked on and their mean click viewpoint, however, exposed disparities between the SERP ranking and SERP display conditions that were overlooked through the lens of clicking diversity. Clicking diversity, as we computed it with the Shannon Index, treats *neutral* as a distinct viewpoint category, rather than as a midpoint on the continuum from attitude-opposing to attitude-supporting. Particularly in the regular SERP display condition, participants primarily clicked on the over-represented viewpoint. In the biased ranking condition, this resulted in a negative mean click viewpoint, while in the neutral ranking condition, the mean click viewpoint was close to zero. We further observed that stance labels without and with explanations, compared to regular search results, reduced the effect of SERP ranking on search behavior and motivated participants to click on search results of the under-represented viewpoints. Thus, supporting suggestions on the benefits of epistemic cues to facilitate complex search tasks [210, 325], the interventions facilitated engagement with diverse viewpoints and countered the effects of biased viewpoint rankings.

3

In addition to the effect of the interventions, we made a noteworthy observation for participants exposed to regular search results. Amongst these participants, clicking diversity is lower for those exposed to a neutral than to a biased ranking. Our investigation of the distinct viewpoints clicked by participants in these groups led us to infer that this disparity might be linked to confirmation bias. While an over-representation of attitude-opposing search results triggers searchers to actively seek under-represented attitude-confirming results, a neutral ranking does not seem to induce the same behavior. To assert whether this behavior is caused by confirmation bias, future research should investigate search behavior in conditions of attitude-supporting bias and whether engagement with the underrepresented viewpoint is observed to a lesser extent when it is attitude-opposing. In fact, participants who were exposed to regular SERPs with neutral rankings were highly unlikely to engage with search results that were either attitude-confirming or attitude-opposing, likely having their information needs satisfied by neutral search results alone. In conditions with stance labels without and with explanations, however, they were motivated to explore search results with diverse labels across ranking conditions.

Findings on Search Effort. By increasing the transparency of the viewpoint ranking we wanted to facilitate engagement with diverse viewpoints. Moreover, we aimed to motivate searchers to invest effort in the search task, for instance, to overcome biased SERP rankings and explore lower-ranked search results and diverse viewpoints. Our findings revealed that stance labels without and with explanations indeed caused participants to invest the effort to engage with under-represented viewpoints in lower-ranked search results. While we did not find consistent differences in the number of clicks and task completion time between the intervention conditions, we observed higher mean task completion times in the condition with explanations. This suggests that some participants invested time and cognitive resources processing the explanations, whereas others did not, aligning with findings on user-dependencies of engagement with and usefulness of explanations [224, 225]. Future research should explore which types of explanations effectively assist different users in detecting incorrect stance labels, and which ones do not yield benefits but merely demand cognitive resources that subsequently cannot be allocated to engaging with diverse viewpoints.

Findings on Attitude Change. Based on our findings, it appears that the search engine manipulation effect as found by Epstein and and Robertson [82], may not play a role for opinionated participants conducting searches on debated topics. This aligns with recent findings on the lesser role of immediate exposure effects for search on social and political topics on which users have pre-existing beliefs [302, 323]. Our findings suggest that the likelihood of attitude change is contingent on the topic, with participants demonstrating a lower inclination to change their attitude after engaging with search results on atheism (affecting the individual) compared to school uniforms (affecting society). Drawing from this observation, we suggest, that future research should investigate whether the degree of consequences of opinions and decisions for the individual may affect the risk of exposure effects, such as the search engine manipulation effect.

Limitations. This user study is not without limitations. We attempted to embody common search environments. Still, the study setup required us to restrict access to the search results in our prepared dataset with stance labels and explanations for each result, which could result in reduced ecological validity. Future research should investigate the effect of stance labels in conditions of unrestricted search access.

The ternary viewpoint categorization and the corresponding definition of clicking diversity treats *neutral* not as a midpoint on a linear scale from attitude-opposing and attitude-confirming, but as a distinct stance category with equal significance to the attitude-opposing and attitude-confirming stances. An additional limitation of ternary stance labels is that they do not sufficiently represent the nuances of viewpoints on debated topics [68]. Nonetheless, we chose ternary stance labels to allow for easily comprehensible stance labels and stance detection models that achieve high accuracy. Yet, this resulted in both the attitude-opposing and neutral ranking equally over-representing one of the three stance categories and consequently in similar observed values of clicking diversity. To address this constraint in the ternary stance labels and the corresponding metric used for calculating clicking diversity, we additionally explored the distinct viewpoints participants clicked on and the mean click viewpoint.

Lastly, we did not include a SERP ranking biased toward the attitude-supporting viewpoint. Future research should investigate whether stance labels encourage searchers to overcome both ranking bias and confirmation bias by engaging with search results of the under-represented viewpoints in such scenarios.

3.6. Conclusions

In this Chapter, we present a user study in which we investigated whether search behavior during web search on debated topics is affected by increased transparency of the viewpoint ranking, which we provided through stance labels that indicate the stance of search results. We found that participants exposed to search results with stance labels engaged with more diverse viewpoints than those exposed to regular search results. Our exploratory insights further indicated that stance labels supported participants in countering exposure effects since they were more likely to click on search results of under-represented viewpoints. We consequently infer that searchers genuinely want to explore diverse viewpoints when provided with tools to differentiate between search results but have become accustomed to relying on opaque search engine rankings, impeding the detection of exposure bias. Drawing from these findings, we posit that epistemic cues such as stance labels, which can reduce the opacity of search engine ranking decisions, could play a pivotal role in shaping the future landscape of information retrieval on debated topics.

II

GUIDING SEARCH BEHAVIOR

In Part II, we focus on mitigating the confirmation bias of strongly opinionated searchers, considering their behavioral patterns that impede responsible opinion formation (e.g., high confirmation bias, low knowledge gain) as detailed in Chapter 2. To mitigate confirmation bias, interventions could aim at *guiding search behavior* towards decreased interactions with attitude-confirming search results. *Nudging interventions* which aim at subtly guiding decision-making without restricting possible choices [49, 337], consisting of *warning labels* and *obfuscations* to reduce the engagement with selected items, have been successfully applied in the context of combating the spread of misinformation [55, 157, 179, 221]. Given the similarity of objectives (i.e., decreased engagement with targeted items), we investigated whether this intervention is likewise successful for confirmation bias mitigation and supporting thorough information-seeking during search on debated topics.

This intervention combines reflective and automatic nudging elements by *prompting reflective choice* with the warning label and *influencing behavior* through the default obfuscation [49]. Yet, automatic nudging strategies have been criticized for being paternalistic, enabling manipulation, and not supporting learning, despite their general success in guiding user behavior [49, 119, 157, 210]. Nevertheless, relying solely on reflective nudging elements could increase the cognitive demand, potentially also increasing searchers' susceptibility to cognitive biases.

Therefore, in II, we focus on the following research question:

> **RQ$_{\text{II}}$**: Can we guide individuals with warning labels and obfuscations to engage in unbiased search behavior on debated topics without harming their autonomy?

In Chapter 4 we addressed **RQ$_{\text{II}}$** with a user study, investigating the effect of warning labels with obfuscations on searchers' clicks on attitude-confirming search results. To understand whether the intervention could be abused to direct search interactions for malicious purposes, we investigated the effect of the intervention applied to randomly selected, in addition to targeted attitude-confirming search results. We found that it reduced interactions with all targeted search results, attitude-confirming or randomly selected, indicating a potential for abuse. Yet, it was unclear whether the warning label, the obfuscation, or both caused the reduced interaction. In Chapter 5, we thus investigated additional exploratory data collected in the first user study, such as users' *cognitive style*—an individual's tendency to rely more on analytic, effortful or intuitive, effortless thinking [46, 90]. Our observations indicated that the extent to which the warning label and the obfuscation were the sources of decreased interaction might vary across users with distinct cognitive styles. To discern which element of the intervention caused the effect of decreased interaction and better understand the role of searchers' cognitive style, we conducted a follow-up study, also described in Chapter 5. We discovered that obfuscations harm user autonomy and run the risk of manipulating search behavior, while the warning label without obfuscation did not exhaust searchers' processing capabilities but encouraged them to choose to click less on attitude-confirming search results.

4

This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias

This chapter is based on a published full conference paper: **Alisa Rieger**, Tim Draws, Mariët Theune, and Nava Tintarev. "This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias". In: *Proceedings of the 32st ACM Conference on Hypertext and Social Media*. Virtual Event USA: ACM, Aug. 2021, pp. 189–199. ISBN: 978-1-4503-8551-0. DOI: 10.1145/3465336.3475101

4

4.1. Introduction

Susceptibility to biases such as confirmation bias has been linked to a lack of analytical thinking, as has susceptibility to misinformation [261]. Given this parallel, our approach to confirmation bias mitigation is inspired by efforts to mitigate the spread of misinformation: research showed that displaying warning labels prior to exposure to misinformation and requiring users' active consent before showing the item is effective in stimulating more skepticism, analytic information processing, and decreasing the interaction with misinformation [157, 170, 198, 221].

We investigated whether showing warning labels prior to exposure to attitude-confirming search results (i.e., search results that support a viewpoint in line with a user's attitude on a topic) could be likewise effective in mitigating confirmation bias during online search. We thus aimed to achieve a decrease in confirmation bias during search by applying search result obfuscations with warning labels. This way, we wanted users to look at a topic from different viewpoints and, consequently, make more informed decisions. This study addresses the following research question:

RQ1: Can search result obfuscations with warnings of confirmation bias mitigate confirmation bias by motivating users to interact with attitude-opposing search results during search for information on debated topics?

We investigated this question by conducting a preregistered, between-subjects user study with crowd-workers.[1] In this study, we observed user interaction (N=328) with search results on four different debated topics, comparing the interaction behavior between six groups (three levels of search result display, and two levels of task).

Our results show that obfuscating search results with warning labels is effective in decreasing interaction with these search results. We also found that targeted obfuscation of attitude-confirming search results causes increased interactions with attitude-opposing search results and thus might be an effective approach to mitigating confirmation bias during search result selection.

The preregistration, data-sets, and material for gathering the data as well as for analyzing the results and replicating our study are publicly available.[2]

4.2. Related Work and Hypotheses

In this section, we look at findings on confirmation bias during search and cognitive bias mitigation. Further, we take a look at approaches to nudge users to a more analytic information processing which were applied to combat online misinformation and at potential user-related factors that might result in behavioral patterns during search. Note that the hypotheses we present here had been preregistered before any collection of data.

[1] Preregistering meant publicly determining our hypotheses, experimental setup, and analysis plan before any data collection. The (time-stamped) preregistration document can be found in our repository: https://osf.io/32wym/.

[2] See link in Footnote 1.

4.2.1. Confirmation Bias During Search

Online search for information has become an indispensable part of our day-to-day life, whether we are trying to settle trivial discussions, looking for information on how best to do something, or collecting information before making the decision on who to vote for in an election. Online information thus affects our decisions, even important life decisions, immensely [50]. To make an efficient decision despite the overwhelming amount of possible choices of search results, we tend to apply search strategies, for example, by searching for information that confirms our prior beliefs [18, 173]. Even though these strategies help us in many cases to make faster and easier decisions by reducing the amount of information and uncertainty [101], they can also do harm when we have the intent to make a well-informed decision but miss out on information supporting another viewpoint than our own. In a broader perspective, such behavior is likely to drive polarization, diminish the quality of public discourse, and contribute to ideological extremism [134, 202]. Regardless of the specific democratic theory one supports, nearly all strands of democratic theory emphasize the importance of promoting viewpoint diversity [242].

When identifying how to mitigate confirmation bias during search, the search process can be divided into three sub-processes during which confirmation bias can occur in different forms: (1) querying for, (2) selecting, and (3) making decisions based on information [162]. We focus on the process of (2) selecting information, during which confirmation bias can be observed in an increased likelihood of interaction (i.e., clicking on or sharing) with search results that confirm our prior beliefs compared to competing possibilities [8, 18]. A widely used measure of confirmation bias during search result selection is thus the number/proportion of selected attitude-confirming search results compared to attitude-opposing ones [162, 172]. In this study, we investigate information selection on two levels (see Section 4.2.3): for oneself by clicking on items (i.e., *clicking behavior*) and for others by sharing items (i.e., *marking behavior*).

4.2.2. Nudges for Confirmation Bias Mitigation

The concept of *nudging* refers to mechanisms that subtly influence users to make decisions that are considered to be beneficial for them, without restricting possible choices [339]. For confirmation bias mitigation during search, nudges could be applied in an indirect approach aiming at generally motivating analytical thinking and supporting users in being more susceptible to genuine evidence, referred to as *nudges for reason* by [196], and in a more direct way aiming at influencing users' item selection behavior and guiding them towards interaction with attitude-opposing search results. This can be achieved by applying nudges that aim to modify the *Decision Structure*; e.g., by ordering items, setting defaults, or altering the required effort [148].

Prior work on confirmation bias mitigation mostly researched approaches of nudging towards a less biased item selection by means of data visualization [106, 201]. Nudges for bias mitigation based on natural language which may generate more immediate transparency for the users [301], have not been studied for confirmation bias mitigation yet. Such nudges have, however, been applied to guide users toward item selection or avoidance for the purpose of combating online misinformation. Previous work on this subject applied warning labels to flag items that may contain misinformation and decreased the ease of access by obfuscating these items by default [170, 198]. This way, users are effectively and transparently nudged towards increased skepticism of and decreased engagement with misinformation [61, 221]. Kaiser et al. [157] found that engagement was further decreased when requiring additional effort such as actively clicking a button. Investigating a similar approach in the context of bias mitigation, Hube et al. [143] found that presenting messages that explicitly make workers aware of potential bias and require interaction to proceed with the task is effective in mitigating bias stemming from worker opinions during crowd-workers labeling tasks. We thus expected that this approach would be likewise effective in mitigating confirmation bias during information selection for oneself (*clicking*) and for others (*marking*) and formulated the following hypotheses:

H1: Users of search engines are less likely to *click* on attitude-confirming search results when some search results on the search engine result page (SERP) are obfuscated with a warning label.

H2: Users of search engines are less likely to *mark* attitude-confirming search results as particularly relevant when some search results are displayed with a warning label.

Next to the search result display, the intention users have when selecting search results is likely to affect their interaction. We will proceed to discuss relevant literature on task design for bias mitigation in the next section.

4.2.3. Effects of Task Design

Cognitive biases are likely to decrease or disappear if task or context stimulate more analytic information processing, for example by triggering high personal accountability or critical thinking in the user [136, 230, 326]. Further, user studies testing a new interface feature such as the one we are presenting here might result in increased interaction with novel features caused by participants' curiosity. This is undesired for this study but has been taken advantage of for nudging users towards certain actions in other studies [139, 343]. Another factor impacting the effectiveness of obfuscations with warning labels is repeated exposure to them. This might lead to initial strengthening and, over a longer term, to habituation and thus weakening of their effect on users' interaction with attitude-confirming search results [12, 289].

To detect potential undesired effects of task design, curiosity, or repeated exposure to the warning label we asked participants to complete two sub-tasks, either in *two separate tasks* or in *one joint task*: (1) explore the SERP as they would do normally and, as a basis of sharing, (2) mark results they considered to be particularly relevant (for a detailed reasoning for this task design see Section 4.3.3). This led us to the following hypothesis:

H3: In the two separate tasks condition, users of search engines are less likely to *mark* attitude-confirming search results as particularly relevant, compared to the joint task condition.

4.2.4. USER-RELATED PATTERNS IN SEARCH BEHAVIOR

In addition to external factors discussed in the previous sections, search result selection can be driven by internal factors which can be both situational or stable and are individually different for different users. Situational factors include factors such as attitude strength, attitude certainty, and interest in the topic. Strong attitudes and high certainty were shown to result in increased confirmation bias [172] and high interest was shown to result in increased information processing capabilities and consequently in more effective information processing [316]. A stable internal factor driving search result selection is, for example, the extent to which users value diverse viewpoints or are challenge averse [8, 239]. Another stable internal factor influencing the reaction to the warning labels we propose for confirmation bias mitigation during search is the susceptibility to persuasive messages [9]. Both factors are closely related to the concept of *Need for Cognition* (NFC). NFC has been described as *The individual's tendency to organize their experience meaningfully [62]* and affects how users interact with information and to which extent this behavior is affected by confirmation bias, and how they process explanations and (persuasive) messages [47, 224, 348].

We thus anticipate that users will have a general propensity to interact with viewpoint-confirming views, or sensitivity to confirmation bias and warning labels and formulate the following hypothesis:

H4: Users are likely to display a consistent pattern of behavior while *clicking* on and *marking* attitude-confirming search results (i.e. participants' marking behavior correlates with their clicking behavior).

4.3. METHOD

To investigate our research questions outlined in Section 4.2, we conducted a between-subjects user study. We manipulated the factors **display** (*targeted obfuscation, random obfuscation, no obfuscation*) and **task** (*two separate tasks, joint task*) and evaluated the degree to which participants would click on and mark attitude-confirming search results.

4.3.1. MATERIALS

TOPICS

Draws et al. [74] provide a dataset containing user attitudes regarding 18 different controversial topics from the website *ProCon* [275]. These 18 topics were selected because the authors assumed that they would be applicable globally and that they would not include highly emotionally charged topics. The authors asked 100 participants to state their attitude towards each of these topics on a seven-point Likert scale ranging from "strongly disagree"(-3) to "strongly agree" (+3). From this dataset, we selected topics for which we expected to observe confirmation bias; i.e., topics where participants reported to have comparatively large proportions of *strong* attitudes. We operationalized this as topics for which at least around 50% of participants selected the options -3, -2, +2, or +3. Following this criterion, four topics were included in the experiment: **(1)** Is Drinking Milk Healthy for Humans?; **(2)** Is

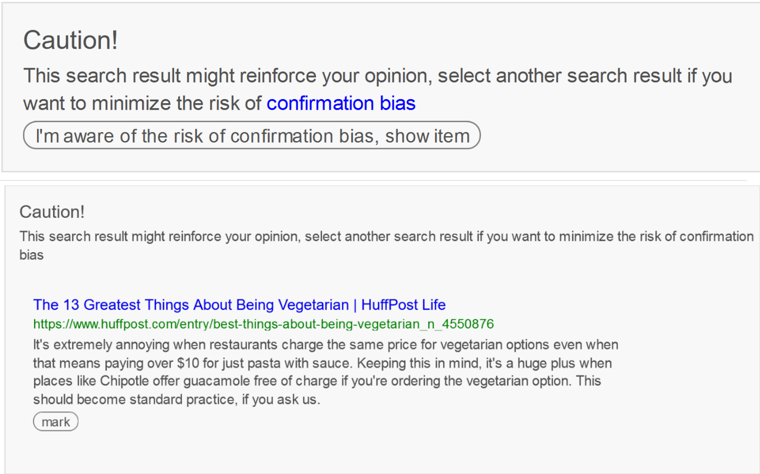


Figure 4.1: Obfuscated search result during SERP exploration task before (top) and after (bottom) *show item* button was clicked. Note, that font size variations are a result of figure formatting; the actual search interface displayed uniform font sizes.

Homework Beneficial?; **(3)** Should People Become Vegetarian?; **(4)** Should Students Have to Wear School Uniforms?

Search Results

Draws et al. [74] provided a dataset of 50 search results for 14 pre-defined queries related to each of the topics using the *Bing API* [223]. From these 700 retrieved URLs per topic, we handpicked 50 opinionated search results by assessing their relevance for each of the four selected topics. The resulting 200 unique search results were subsequently annotated by crowd-workers on *Amazon Mechanical Turk* [350]. Specifically, workers annotated the relevance to the topic (binary) and the viewpoint with respect to the topic (scored on a seven-point Likert scale ranging from "strongly opposing" to "strongly supporting"). We collected three annotations for each search result and observed a satisfactory inter-annotator agreement (*Krippendorff's* $\alpha = 0.78$) [126]. In our final dataset, the search result was assigned the median value of these three annotations. Per topic, we subsequently selected 12 search results by randomly sampling two "strongly supporting", two "supporting", two "somewhat supporting", two "somewhat opposing", two "opposing", and two "strongly opposing" from all search results that were deemed relevant by all crowd-workers.[3] They were displayed in random order (see Table 4.1).

Search Result Obfuscation

Search results were obfuscated with a warning label, warning of the risk of confirmation bias if this item is selected and advising the participant to select

[3]Datasets containing the 12 included search results per topic as well as all 200 annotated search results are publicly available at link in Footnote 1.

Table 4.1: Representation of the search result display. Each row represents a search result (twelve in total), with two results per viewpoint (-3 to +3), displayed in random order (example); Obfuscation illustrated with [[ ]]. Participants in the targeted obfuscation condition would see either targeted obfuscation supporting or opposing, depending on their initial attitude toward the assigned topic.

No Obfuscation	Targeted Obfuscation sup	Targeted Obfuscation opp	Random Obfuscation
-1	-1	-1	[[-1]]
+2	[[+2]]	+2	[[+2]]
-3	-3	[[-3]]	-3
-2	-2	[[-2]]	-2
+3	[[+3]]	+3	+3
-1	-1	-1	-1
+2	[[+2]]	+2	[[+2]]
+1	+1	+1	+1
-2	-2	[[-2]]	-2
+3	[[+3]]	+3	+3
+1	+1	+1	+1
-3	-3	[[-3]]	[[-3]]

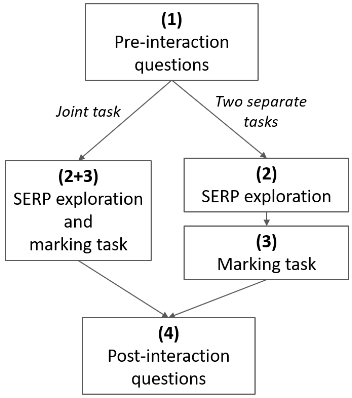


Figure 4.2: Data collection procedure for *joint task* and *two separate tasks* condition: pre-interaction questions, SERP exploration, marking task, and post-interaction questions.

another item (see the left-hand panel in Figure 4.1). Here, the *Wikipedia* entry on confirmation bias [378] was linked so that participants could inform themselves about this cognitive bias. To view the obfuscated search result, participants had to click a button, stating they were aware of the risk of confirmation bias. After clicking the button, the search result would become visible underneath the warning label (see the right-hand panel in Figure 4.1). During the marking task in the *two separate tasks* condition, obfuscated search results were displayed in the same way (search result visible below warning label).

4.3.2. VARIABLES

Independent variables

- *Display* (categorical, between-subjects). Participants were randomly assigned to one of three display conditions: (1) targeted obfuscation of moderate and extreme attitude-confirming search results (see Figure 4.1), (2) obfuscation of four randomly chosen search results, and (3) no obfuscation (see Table 4.1).

- *Task* (categorical, between-subjects). Participants were also randomly assigned to one of the two task conditions: (1) a *two separate tasks condition*, where search result exploration and marking particularly relevant results was split into two separate tasks or (2) a single *joint task condition*, where search result exploration and marking particularly relevant results were done together (see Figure 4.2).

Dependent variables

- *Click proportion attitude-confirming results* (continuous). Proportion of attitude-confirming results among the search results that participants clicked on during search results exploration: [0,1].

- *Marking proportion attitude-confirming results* (continuous). Proportion of attitude-confirming results among the search results that participants marked when asked for items they would share: [0,1].

Exploratory variables

- *Click proportion obfuscated search results.* For targeted and random obfuscation condition: proportion of obfuscated results among the search results participants clicked on during search results exploration.

- *Marking proportion obfuscated search results.* For targeted and random obfuscation condition: proportion of obfuscated results among the search results participants marked when asked for items they would share.

Descriptive variables

- *Gender.* Participants could select between “female”, “male”, or “non-binary/other”.

- *Age.* Participants were asked to enter their age using a numerical value.

- *Time spent on the task.* Time participants spent on the whole task, including prior- and post-interaction questions.

- *Time spent on SERP exploration.* Time participants spent on the SERP exploring the search results.

- *Number of clicks.* Number of search results participants clicked on to retrieve the linked document.

- *Number of markings.* Number of search results participants marked as being particularly relevant.

4.3.3. Procedure

We conducted this study on the online survey platform *Qualtrics* [283]. To control for data quality, we integrated four attention checks into the survey (two prior to the clicking and marking task and two during post-interaction questions). In these checks, we explicitly instructed participants on which specific response to provide (e.g., *This is an attention check. Please select 'Agree'.*). The procedure, approved by the ethics committee of our institution, consisted of four subsequent steps (see Figure 4.2):

Step 1: Pre-interaction. Participants were given a short introduction to the experiment and asked to answer demographic questions (gender, age). We then asked them to imagine the following scenario: *“You had a discussion with a relative or friend on a certain topic. The discussion made you curious about the topic, and to* *inform yourself further, you are conducting a web search on the topic."* Subsequently, we asked participants to state their attitude towards the four selected topics (see Section 4.3.1) on a seven-point Likert scale ranging from "strongly disagree" to "strongly agree". The responses "strongly disagree", "disagree", "agree", and "strongly agree" were considered to be *strong attitudes.*

Step 2 and 3: Search result exploration and marking. Based on their answers during Step 1, participants were randomly assigned to (1) one of the topics they held a strong attitude towards,[4] (2) a **display** condition, and (3) a **task** condition. They were thus exposed to a randomly ordered list of search results relevant to their assigned topic in their assigned search result display format. Participants' task was to explore the search results (Step 2) and mark search results that they considered to be particularly relevant (Step 3). Depending on their assigned task condition, they would perform these actions separately or together:

- *Two separate tasks condition.* Participants saw the list of 12 search results (with obfuscations depending on their assigned **display** condition) relevant to their assigned topic (*SERP exploration*). They were given as much time as they wanted to explore the search results and examine the linked documents.[5] After continuing to the next page, participants were again presented with all 12 search results. Among those results, they were then asked to mark items that they considered to be particularly relevant and informative and that they would have liked to forward to a relative who wants to form an opinion on the topic (*marking task*). Search results that were obfuscated during SERP exploration were still displayed with the warning but not obfuscated (see Figure 4.1). Participants were not able to examine the linked documents again but could only see the titles and snippets.

- *Joint task condition.* As above, but in this condition, participants could mark items that they considered to be particularly relevant and informative (*marking task*) at the same time as they explored the search results (*SERP exploration*).

Step 4: Post-interaction. We asked participants to state their attitude on the selected topic again and to answer a number of questions on their experience with the task, self-perception of their behavior, and user-experience.[6]

Reasoning for task design. To be able to draw valid conclusions from the collected data, we attempted to design a task and scenario that would motivate participants to mimic their natural search result exploration behavior by requiring a low level of accountability. However, since the feature of obfuscations with warning labels was novel, we had to control for the potential effects of curiosity. We did so by observing a second level of interaction behavior, for which we expected users to be less driven

[4]Participants who did not hold a strong attitude towards any of the four topics were ejected from the study; see Section 4.3.4.

[5]We intentionally left the duration for exploration up to the participants to best mimic natural exploration behavior.

[6]Further details on the post-interaction questionnaire are available at link in Footnote 1.

by curiosity because this second task required increased accountability. We thus observed two levels of participant behavior, first (1) exploring search results for themselves *(clicking: low accountability, potentially high curiosity)*, and then (2) for others *(marking: high accountability, low curiosity)*. However, by asking participants to mark particularly relevant search results in a separate task and displaying the warning labels again, we could have introduced an unwanted effect of repeated exposure to the warning labels. To allow us to single out potential undesired effects of task, curiosity, or repeated exposure to the warning label we asked participants to complete the two tasks, either in *two separate tasks* or in *one joint task*.

4.3.4. Sample

4

Before collecting data, an a priori power analysis for a between-subjects ANOVA (with $f = 0.25$, $\alpha = \frac{0.05}{4} = 0.0125$ (i.e., due to testing four different hypotheses), and $(1-\beta) = 0.8$) determined a required sample size of 282 participants.

We initially recruited a total of 510 participants via the online participant recruitment platform *Prolific* [277]. Participants were required to be at least 18 years old and to speak English fluently. They were allowed to participate only once and were paid £1.75 for their participation (*mean* = £7.21/h), following the prolific recommendations for fair pay at the time of data collection. From these 510 participants, 182 were excluded from data analysis because they did not fulfill the inclusion criteria: they did not report having a strong attitude on any of the topics *(41)*, failed at one or more of four attention checks *(50)*, spent less than 60 seconds on the SERP (*80*), or did not click on and mark any search results *(11)*.

Of the remaining 328 participants (gender: 49% female, 51% male, <1% non-binary/other; age: $mean = 28.8, sd = 10.6$), 282 clicked on at least one search result and thus were included in testing **H1** (clicking behavior), 293 marked at least one search result and thus were included in testing **H2** (marking behavior) and **H3** (task difference marking behavior), and 248 clicked on *and* marked at least one search result and thus were included in testing **H4** (correlation clicking and marking behavior). Participants were randomly assigned to one of the topics for which they reported to strongly agree or disagree (-3, +3). If they did not report strongly agreeing or disagreeing with any of the four topics, they were assigned a topic for which they reported agreeing or disagreeing (-2, +2). If participants did not report a strong attitude (-3, -2, +2, +3) on any of the four topics, they were not able to participate further but received partial payment (£0.50).

4.3.5. Statistical Analysis

To test our hypotheses we planned to apply a one-way ANOVA to compare the clicking behavior between the three **display** conditions, and a two-way ANOVA to compare the marking behavior between the three **display** and the two **task** conditions. The terms clicking and marking behavior refer to the proportion of clicks on and markings of attitude-confirming search results. However, *Shapiro-Wilk* tests revealed that our observations were not normally distributed. Hence, we applied *Kruskal-Wallis* tests for testing **H1** (clicking behavior), **H2** (marking behavior), and **H3** (task difference marking behavior). For pairwise post-hoc testing of differences between **display** conditions, we applied *Dunn* tests. To test **H4** (relation clicking and marking behavior), we conducted a *Spearman's* rank correlation analysis. For testing all four hypotheses, the significance threshold was set at $\alpha = \frac{0.05}{4} = 0.0125$, aiming at a type 1 error probability of $\alpha = 0.05$ and applying Bonferroni correction to correct for multiple testing. For the post-hoc *Dunn* tests for **H1** and **H2** of differences between the three **display** conditions, the significance threshold was set to $\alpha = \frac{0.05}{3} = 0.0167$ each, due to testing 3 pairwise comparisons. All analyses were conducted in *R* [284].

4.4. RESULTS

In the following section, we will present the collected data by means of descriptive statistics and the results of hypotheses testing as specified in Section 4.3.5.

4

4.4.1. DESCRIPTIVE STATISTICS

Participants' distribution over the 6 different conditions (three display and two task conditions) was comparable: 49 to 66 participants completed the task in each condition. The criterion for topic assignment resulted in 64, 94, 68, and 102 participants being assigned to topics 1, 2, 3, and 4, respectively. The average time spent on the task was 17.3 minutes ($se = 0.5$) with no difference between conditions. The time spent on the SERP page was 4.8 minutes for the **joint task** condition ($se = 0.3$) and 4.1 minutes for the **two separate tasks** condition ($se = 0.3$) with no differences between **display** conditions. We recall that 80 participants were excluded from the study for spending fewer than 60 seconds on the SERP, but note that these durations are substantially higher than 60 seconds.

With regards to the level of interaction, the mean number of clicks during search result exploration was 3.02 for the **joint task** condition ($se = 0.2$) and 2.57 for the **two separate tasks** condition ($se = 0.15$) with no differences between **display** conditions. This reflects roughly 3/12, or 25% of the search results. The mean number of markings was 2.95 ($se = 0.11$, no difference between conditions). This degree of interaction is consistent with the qualitative feedback, which suggests that participants understood the task well and found it interesting and enjoyable.

4.4.2. HYPOTHESES TESTING

H1 - OBFUSCATIONS WITH WARNING LABELS RESULT IN A LOWER PROPORTION OF CLICKS ON ATTITUDE-CONFIRMING SEARCH RESULTS.

The results of a Kruskal-Wallis test for the click behavior show evidence for a moderate effect of search result **display** on the proportion of attitude-confirming clicks ($H(2) = 33.87, p < .001$, $\eta^2 = .11$). A pairwise post-hoc Dunn test shows that the proportion of clicks on attitude-confirming search results was significantly lower in **targeted obfuscation** ($mean = 0.34, se = 0.03$) compared to **random obfuscation** ($mean = 0.54, SE = 0.03$; $p < .001$) and **no obfuscation** ($mean = 0.58, SE = 0.03$; $p < .001$; see Figure 4.3). However, there was no difference in the clicking behavior between the **random obfuscation** and **no obfuscation** conditions, leaving our hypothesis only partially confirmed.

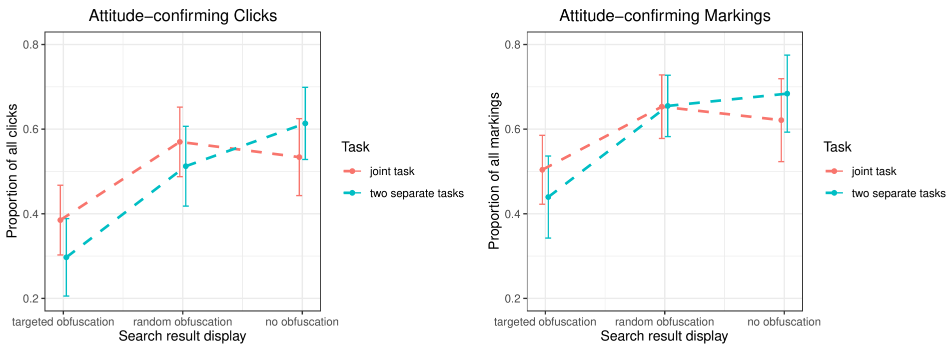


Figure 4.3: Interaction with attitude-confirming search results: mean proportion of participant's ***(H1)*** attitude-confirming clicks (left) and ***(H2)*** markings (right) per **display** condition (targeted obfuscation, random obfuscation no obfuscation) and per ***(H3)*** **task** condition (joint task and two separate tasks) with 95% confidence intervals. A proportion of one implies that all of a participant's clicks/markings were on attitude-confirming search results.

H2 - OBFUSCATIONS WITH WARNING LABELS RESULT IN A LOWER PROPORTION OF MARKINGS OF ATTITUDE-CONFIRMING SEARCH RESULTS.

The results of a Kruskal-Wallis test for the marking behavior likewise show evidence for a moderate effect of the factor **display** on the proportion of attitude-confirming markings ($H(2) = 21.23, p < .001, \eta^2 = .07$). A pairwise post-hoc Dunn test shows that the proportion of markings of attitude-confirming search results was significantly lower in **targeted obfuscation** ($mean = 0.47, SE = 0.03$) compared to **random obfuscation** ($mean = 0.65, SE = 0.03$; $p < .001$) and **no obfuscation** ($mean = 0.66, SE = 0.03$; $p < .001$; see Figure 4.3). As was the case for the clicking behavior, there was no difference in the marking behavior between **random obfuscation** and **no obfuscation**.

H3 - TWO SEPARATE TASKS CONDITION RESULTS IN A LOWER PROPORTION OF MARKINGS OF ATTITUDE-CONFIRMING SEARCH RESULTS.

Against our hypothesis, the result of a Kruskal-Wallis test for the marking behavior does not show evidence for an effect of the factor **task** on the proportion of attitude-confirming search results ($H(2) = 0.04, p = .83$).

H4 - BEHAVIORAL PATTERN: CLICKING AND MARKING BEHAVIOR ARE CORRELATED.

A Spearman rank correlation test shows evidence for a substantial positive correlation between the proportion of attitude-confirming clicks and markings ($\rho = .51, p < .001, R^2 = .26$). After controlling for the effect of **display** on the relationship, we still found clicking and marking behavior to be moderately positively correlated ($\rho = .44, p < .001, R^2 = .2$). This finding supports our hypothesis that participants would be displaying a consistent pattern of behavior across both tasks.

4.4.3. EXPLORATORY RESULTS

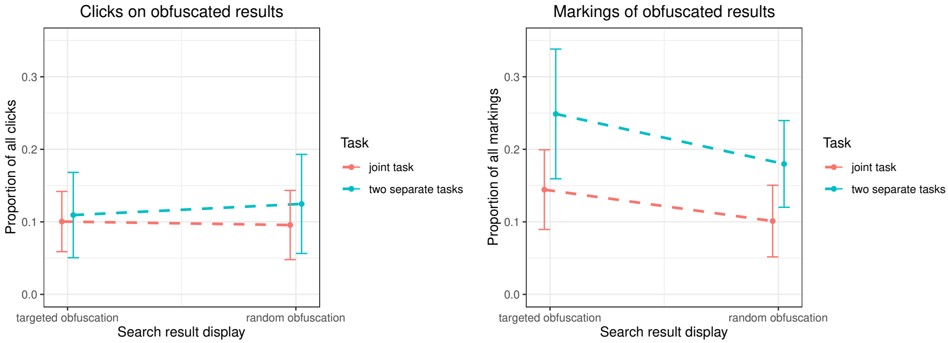


Figure 4.4: Interaction with obfuscated search results: mean proportion of participant's clicks on (left) and markings of (right) obfuscated search results in the **targeted obfuscation** and **random obfuscation** condition and per task condition (joint task and two separate tasks) with 95% confidence intervals. A proportion of 0.1 implies that 10% of a participant's clicks/markings were on obfuscated search results.

INTERACTION WITH OBFUSCATED SEARCH RESULTS.

While obfuscated search results in the **targeted obfuscation** and the **random obfuscation** conditions make up a proportion of 33% of all displayed search results, the mean proportion of clicks on these search results is only 10% (see Figure 4.4). We observed no difference in this proportion of clicks between **targeted obfuscation** and **random obfuscation** conditions. For the marking behavior, this mean proportion is similar for the **joint task** condition ($mean = 0.12, SE = 0.02$), but higher for the **two separate tasks** condition ($mean = 0.21, SE = 0.03$).

4.5. DISCUSSION

In this study, we investigated whether item obfuscations with warning labels would be effective in countering confirmation bias in web search. We conducted a between-subject user study for which we manipulated the factors **display** *(targeted obfuscation, random obfuscation, no obfuscation)* and **task** *(two separate tasks, joint task)*. We evaluated in which proportion the participants would click on and mark attitude-confirming search results.

Effect of display. While we found that targeted obfuscations with warning labels decreased the likelihood of interacting with (clicking, marking) attitude-confirming search results compared to no obfuscation, we did not find any effect of the random obfuscation condition. This finding implies that the mere presence of warning labels does not motivate users to decrease interaction with attitude-confirming search results, but that targeted obfuscations are required to achieve this.

However, when looking at the interaction with obfuscated instead of attitude-confirming search results, we found that both targeted and random obfuscations of search results were effective in decreasing the proportion of clicks on these obfuscated search results. This implies that search result obfuscations are a powerful tool in steering users' search result selection behavior, which consequently could be misused for purposes other than the users' benefit, raising ethical considerations which we follow up on in Section 4.5.1. This observation could be explained in two ways: either (1) participants blindly trusted the system's decision, hindering them from realizing that in the random obfuscation condition, the results obfuscated were indeed random (and not attitude-confirming), or (2) they simply ignored the obfuscated search results and focused on clearly visible search results that did not require additional effort. In the targeted obfuscation condition these were to a high proportion (75%) attitude-opposing. The second explanation is in line with the findings of Kaiser et al. [157] that warnings are more effective in decreasing interaction with an item when they require user interaction, partly due to the additional effort introduced to the users' workflow. They warned that this might decrease user experience, not foster informed decision-making, and result in habituation effects.

Further, we consider our findings in light of the *Elaboration Likelihood Model* (ELM) [262], which distinguishes between a *central* and a *peripheral* route to persuasion. It seems likely that the effect we observed was caused by the *peripheral* route (i.e., interacting with attitude-opposing search results because interaction requires less effort and is thus more attractive). The authors stated that attitude change caused by peripheral instead of central cues is less enduring, relatively temporary, and unpredictive of behavior. An approach applying central cues was investigated by Hube et al. [143]. They found that in a setting with no option of choosing a path of lower effort, warnings that require user interaction were effective in mitigating worker bias and improving performance. Thus, for a setting in which users can choose a path of lower effort, we should strive to find an effective combination of peripheral cues to guide users' interaction and to catch their attention and central cues to motivate careful and analytic consideration of information.

Effect of task. We did not find any evidence for an effect of task on the proportion of attitude-confirming markings. This implies that repeated exposure to the warning label (two separate tasks), does not alter the effect of obfuscations with warning labels on markings of attitude-confirming search results, at least to the limited extent to which we were able to observe this in a single session experiment. However, we observed that participants in the joint task condition tended to spend a longer time on the SERP and to click on more search results than participants in the two separate tasks condition. While the former observation might be explained by participants in the joint task condition doing two tasks (clicking and marking) instead of one (only clicking), the latter suggests that the marking task might have motivated increased clicking on search results. Further research on the potential effect of task design on confirmation bias and analytical information processing is required.

Exploratory: Interaction of task and display. During exploratory analysis, we furthermore observed a higher proportion of markings of *obfuscated* search results in the two separate task condition than in the joint task condition for targeted and random obfuscation. This observation might be explained by a task design decision for the separate task condition (see Section 4.5.3): recall that in the two separate tasks condition, during the marking task, search results were not obfuscated but merely displayed with the warning label (see Section 4.3.1). In the joint task condition, however, clicking and marking were done in a single task, and obfuscated search results remained obfuscated unless participants actively clicked the button to view the search result. Removing obfuscation (two separate tasks) seemed to result in more interaction, suggesting that obfuscation is effective in decreasing interaction with search results. This observation further supports the explanation that participants might have chosen the path of lowest effort [157] due to peripheral cues of persuasion for behavioral change [262] instead of carefully and analytically considering the information. Thus, follow-up research could be tailored specifically to answer the question of what causes decreased interaction with search results that are obfuscated with a warning label.

Effect of user-related behavioral patterns. We found evidence for a correlation between participants' proportion of attitude-confirming clicks and markings. This finding suggests that behavioral patterns caused by user-related factors that influence the interaction with search results of different viewpoints exist. Hence, further research is required to investigate how situational or stable factors that have been found to moderate users' search behavior, such as attitude strength [172], interest in the topic [316], and personality traits (e.g. Need for Cognition [348]), might affect the effectiveness of confirmation bias mitigation approaches and how they should be adapted accordingly.

Considerations for real-world applications. Collecting viewpoint annotations for a handpicked selection of 200 search results and specifically asking participants for their attitude on the four selected topics was necessary for conducting our controlled user study but limits the applicability of our approach to complex real-world scenarios. Enabling effective real-world applications of confirmation bias mitigation strategies in search may, however, be possible by drawing from related research. For instance, recent advances in automatic stance detection [5, 184] and perspective discovery [69] provide important tools for assigning correct viewpoint labels automatically. Furthermore, approaches to automatically measure viewpoint diversity in search results [73] or real-time confirmation bias detection, as researched, for instance, for the field of visual analytics [367], might prove useful here. Real-world application of the confirmation bias mitigation approach investigated with this work will lend itself more to large-scale implementation as such tools become more advanced. However, our findings urge us to exercise caution when going about a real-world implementation of such approaches due to a number of ethical considerations which we discuss in the following section.

4.5.1. Ethical Considerations

The two potential explanations for no differences in the interaction behavior with obfuscated search results between targeted and random obfuscation conditions raise ethical concerns with regard to using obfuscations with warnings for confirmation bias mitigation during web search: (1) If the findings can be explained by high trust in warning labels (i.e., even if they are applied incorrectly as was the case in the random obfuscation condition), this would allow for exploitation and misuse for someone's interests and against the user's benefit. (2) If, on the other hand, the findings can be explained by the users' (potentially unintentional) ignorance of or blindness to obfuscated search results and their tendency not to engage with search results if engagement requires additional effort (i.e., clicking one button), then we would be battling cognitive bias by harnessing other cognitive biases. This is effective in getting users to interact with attitude-opposing search results but most likely is not an appropriate approach to motivate analytic information processing. Consequently, this approach would threaten user autonomy and thus not fulfill the requirements for *nudges to reason* stated by [196]. An improvement could be to design obfuscations with warning labels more saliently so that it is less likely that users unintentionally ignore them. However, this proposal requires further research.

Additional ethical considerations concern the practical implementation of approaches of confirmation bias mitigation during search. As discussed in the previous section, users' attitudes on a topic would have to be elicited automatically from their search behavior. However, automatically eliciting personal attitudes on different topics, including sensitive personal information such as political beliefs, requires user-data collection and storage that is not compatible with GDPR regulations. Thus, we promote approaches that base the decision on what to obfuscate merely on the observed behavior in a single search session. This could be done, for instance, by applying targeted obfuscations after a user has selected a number of articles all supporting the same viewpoint. Approaches of real-time confirmation bias detection during search, which do not require storing sensitive user data, need to be examined further in future studies.

Based on these considerations, we propose the following **ethical guidelines**:

1. Apply obfuscations for confirmation bias mitigation exclusively to the users' benefit.

2. Obtain users' consent before obfuscating to mitigate confirmation bias and enable consent withdrawal.

3. Explain transparently why an item is obfuscated so that users understand the system's decision and are able to detect system errors (i.e. incorrect obfuscations).

4. Include simple mechanisms that allow users to control/correct the obfuscation feature if necessary due to incorrect system decisions or as desired by the user.

4.5.2. Implications and Design guidelines

From our findings we learn that obfuscations with warning labels, requiring additional effort to view a search result, are an effective approach to decrease interaction with search results, whereas our exploratory findings suggest that the search result obfuscation might have had a greater impact than the warning label. If such obfuscations are applied targeting attitude-confirming search results, they can effectively nudge users to interact with a higher proportion of attitude-opposing search results than they would do without obfuscations. Thus, if applied carefully, this approach might help users to overcome confirmation bias while selecting search results during online search.

Our findings have practical implications for the implementation of obfuscation-based approaches for confirmation bias mitigation during search, thus we formulated the following **design guidelines**:

1. Design obfuscation in a way that requires an appropriate amount of additional effort to view the item (i.e. button to actively accept the risk of confirmation bias). *Given our findings it seems that users are likely to take the path of lowest effort and thus interact less with items that would require additional effort. However, according to Kaiser et al. [157] the amount needs to be selected diligently to avoid decreasing user experience and warning fatigue.*

2. Select diligently what to obfuscate (target attitude-confirming search results). *Our findings show that users interact less with obfuscated search results no matter if obfuscation is targeted or random. Thus which items to obfuscate should be decided carefully and, if necessary, be adapted.*

3. Design obfuscations with warning labels with an appropriate level of salience to avoid unintentional ignorance of or blindness to the obfuscated items, which would threaten user autonomy. *Our research design failed to detect if the cause of less interaction with obfuscated search results might have been users' ignorance of, or blindness for these search results. However, to fulfill the requirements for ethically permissible nudges that do not threaten user autonomy stated by [196], it is important to ensure that obfuscated search results are not unintentionally ignored because they did not capture users' attention.*

4

4.5.3. Limitations and future work

To be able to conduct a controlled user study, we had to construct an artificial scenario, for which we pre-selected the topics and search results. Even though we assigned each participant to one of the topics for which they reported to have the strongest attitude, they still might not have had a great interest in the topic, or, as formulated by [18], they had "no skin in the game". Further, the setup did not allow participants to formulate one or multiple queries and to conduct the search, as they would naturally do, but they were forced to interact exclusively with the 12 pre-selected search results. However, we attempted to make the task as realistic and relatable as possible and refrained from enforcing minimum time requirements, even though this meant excluding data.

Another limitation of this study is that we only observed one single search session, exposing participants to obfuscations with warning labels for a limited time. Yet, most of us use search engines multiple times per day and thus are exposed to the search engine interface very frequently and adapt our behavior according to our intentions and the search engine's features. It would, therefore, be interesting to observe user behavior and potential adaptions to warning labels and obfuscations in a less controlled and more natural setting and over a longer period of time.

Lastly, we did not investigate the effects of warning labels and obfuscations independently and systemically. We did, however, display search results merely with a warning label (not obfuscated) during the marking task for the two separate tasks condition and compared the interaction behavior to the joint task condition, in which search results were obfuscated with a warning label. Yet, this decision constitutes another limitation of this study since it was based on attempting to control for and investigate the effects of multiple exposures to both the warning labels and the search results, as discussed previously. Ultimately, this design decision prevents us from drawing valid conclusions on the effect of multiple exposures to the warning labels but unveils that obfuscations with warning labels might have been more effective than merely warning labels in decreasing interaction with search results. Targeted investigation of the effects of warning labels and obfuscations independently needs to be done in future studies.

4.6. Conclusions

We presented a user study investigating the effect of obfuscations with warning labels about confirmation bias, on the interaction with viewpoint-annotated search results on debated topics. We found that obfuscations result in decreased interaction with search results and that targeted obfuscations of attitude-confirming search results are effective in increasing the interaction with attitude-opposing search results. However, it remains to be clarified whether this effect was observed because participants trusted the warning label or avoided additional effort and ignored the obfuscated search results. Given these findings, we call for strict regulations, allowing an application of search result obfuscations exclusively to the users' benefit, with their consent, and in a transparent and controllable way.

5

Nudges to Mitigate Confirmation Bias during Web Search on Debated Topics: Support vs. Manipulation

This chapter is based on a published journal paper: **Alisa Rieger**, Tim Draws, Mariët Theune, and Nava Tintarev. "Nudges to Mitigate Confirmation Bias during Web Search on Debated Topics: Support vs. Manipulation". In: *ACM Transactions on the Web* (2023). DOI: 10.1145/3635034

5.1. Introduction

Warning labels and obfuscations to mitigate confirmation bias during web search on debated topics combines both transparent reflective and transparent automatic nudging elements [119] (see Figure 5.1): it *prompts reflective choice* by presenting warning labels and it *influences behavior* by decreasing the ease of access to the item through default obfuscations [49]. To understand the benefits and risks of warning labels and obfuscations to mitigate confirmation bias during search on debated topics in strongly opinionated searchers (see Figure 5.2), we conducted two user studies (see Figure 5.3).

With the first, which is partially presented in Chapter 4, we investigated the effect of warning labels and obfuscations combined (**warning label and obfuscation study**) on searchers' confirmation bias. In Section 5.3, we provide a reiteration of the key findings from Chapter 4, as well as an extended analysis that covers previously uninvestigated exploratory data on attitude change, awareness of bias, and the role of cognitive style. *Cognitive style* describes an individual's tendency to rely more on analytic, effortful or intuitive, effortless thinking [46, 90]. The **warning label and obfuscation study** revealed that warning labels and obfuscations effectively reduce interaction with targeted search results ($f = 0.35$). Yet, it is unclear what specifically caused the observed effect; i.e., whether **(1)** participants read the warning label (reflective nudging element) and, now aware of confirmation bias, actively decided to interact less with attitude confirming search results, or **(2)** participants took the path of lowest effort and unconsciously ignored all obfuscated items (automatic nudging element), since interaction with those required increased effort. Our exploratory findings suggest that the extent to which the reflective or automatic elements of the intervention caused the effect might vary across users with distinct cognitive styles. Further, the exploratory observations suggest that both the search result display and the individuals' cognitive style might impact the searcher beyond their search interactions, namely their attitude change and awareness of bias.

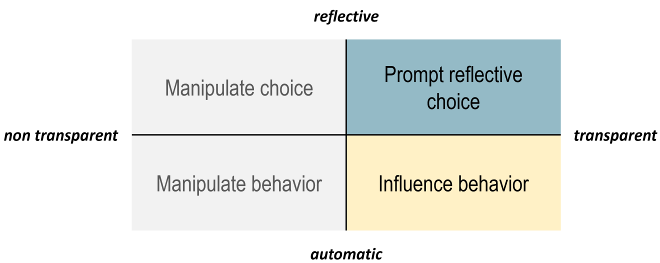


Figure 5.1: Categories of nudging elements, adapted from Hansen and Jespersen [119]. Since this work investigates interventions that aim at guiding (as opposed to manipulating) user behavior, we only consider nudges from the transparent categories.

To better understand what caused the effect of decreased interaction and how the interventions impact searchers with distinct cognitive styles, we initiated the **automatic vs. reflective study** as a follow-up study. With the **automatic vs. reflective**

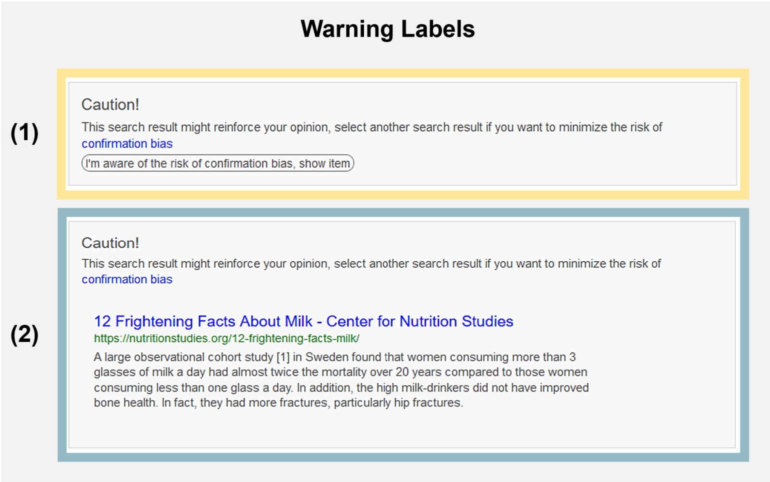


Figure 5.2: Warning labels. (1) Warning label with obfuscation, after participants clicked on *show-button*, the search result was revealed and they saw (2) Warning label without obfuscation. In the *warning label without obfuscation* conditions in the automatic vs. reflective study, the default shown to participants was condition (2).

study detailed in Section 5.4, we tested the effect of the reflective element of the intervention separately by adding a search result display condition (for an overview of the conditions, see Figure 5.3) in which search results were displayed with the warning label (reflective) but not obfuscated (see (2) in Figure 5.2). With this second study, we addressed the following research questions:

RQ1: Can warning labels with and without obfuscations reduce clicks on attitude-confirming search results?

RQ2: Are participants' clicks on search results with warning labels affected by the search result display (with or without obfuscations) and is this potential difference moderated by their cognitive style?

RQ3: Is participants' attitude change affected by the search result display, and/or their cognitive style?

RQ4: Is participants' awareness of bias affected by the search result display, and/or their cognitive style?

The **automatic vs. reflective study** replicated the finding of a moderate effect of obfuscations with warning labels that reduced clicks on attitude-confirming search results for a new set of search results ($f = 0.30$). Moreover, we observed that warning labels without obfuscation (reflective) reduce engagement when applied to attitude-confirming search results, but, in contrast to warning labels with obfuscation, do not reduce engagement when applied to randomly selected search results. Thus, our key takeaways from both studies are that obfuscations, and

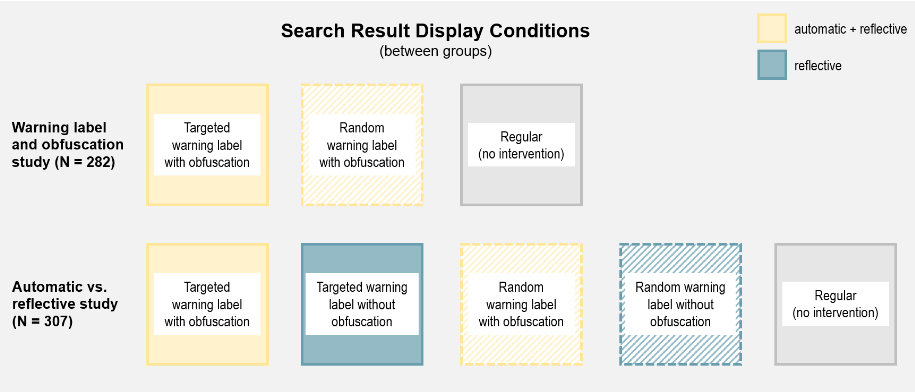


Figure 5.3: Search result display conditions in the warning label and obfuscation study (top) and automatic vs. reflective study (bottom).

possibly other automatic nudging elements, run the risk of manipulating behavior instead of guiding it while warning labels without obfuscations effectively encourage users to choose to engage with less attitude-confirming search results.

5.2. Related Work

In this section, we discuss background literature on different related areas of research. These include search on debated topics and confirmation bias, interventions to guide web interactions, and the role of cognitive reflection during engagement with information.

5.2.1. Search on Debated Topics and Confirmation Bias

Individuals may turn to web search to develop or revise their opinions on different subject matters, e.g., to satisfy individual interest or to gather advice before making decisions [50, 370]. This can concern *debated topics*, subjects on which individuals or groups have different opinions, for instance, due to conflicting values, competing interests, and various possible perspectives from which to view the issues. Web search on debated topics can be consequential for both individuals and society at large, given its potential to influence practical decision-making [50, 82, 219]. Thus, we are interested in how web search could support people in forming opinions responsibly.

The notion of *responsibility* in opinion formation has been thoroughly discussed by philosophers in the field of epistemology [177, 260]. Kornblith [177], for instance, reasons that responsible beliefs are the product of actively gathering evidence and critically evaluating it. For responsible opinion formation, individuals should thus gather information to gain a well-rounded understanding of the topic and the various arguments, and form opinions and make decisions based on the synthesized information they gathered and knew before. Traditionally, the objective of gaining a well-rounded understanding of the topic and arguments could be supported by (public) media and news outlets that are subject to regulations and ethical guidelines, e.g., regarding quality and diversity of content [130]. However, rather than primarily consulting curated journalistic content, people increasingly rely on search engines to actively search for information on debated topics to form opinions or make decisions [50, 370]. The opaque nature of search engines that automatically filter and rank resources and are not (yet) bound to follow principles of responsible information proliferation (e.g., exposure diversity [130]), can prevent users from recognizing whether the provided information is complete and reliable [226, 325]. Web search for responsible opinion formation thus requires self-reliant, thorough, exploratory search behavior, which is known to be cognitively demanding [134, 263, 271].

As a means of simplifying complex search tasks, searchers are prone to resort to heuristics and systematic shortcuts [18]. While such shortcuts typically lead to more efficient actions and decisions under constraint resources (e.g., information-processing capacities or time) [102], they can result in *cognitive biases*—systematic deviations in judgment and decision-making [351]. A prevailing strategy to limit the cognitive demand of search tasks is the *confirmation bias*—the human tendency to prioritize information that confirms prior attitudes [244]. Confirmation bias thus impedes engagement with diverse viewpoints and can manifest throughout the various stages of the information search procedure: it can cause users to employ affirmative testing techniques while querying, interact mainly with search results that align with their attitudes, and disregard information that counters their attitude when evaluating arguments to form beliefs or make decisions [18, 360, 372, 383]. Yet, search engines could be designed to accommodate more complex and exploratory search tasks and support thorough and unbiased information-seeking strategies [320, 325].

5.2.2. Guiding Web Interactions

To empower individuals online, Lorenz-Spreen et al. [210] propose *effective web governance* through the application of behavioral interventions to improve decision-making in a web context, e.g., by applying *nudges*. Nudges are interventions that subtly guide users to make better decisions without restricting possible choices, e.g., by setting defaults, creating friction and altering the required effort, or suggesting alternatives [49, 337].

Caraban et al. [49], grouped different nudging approaches according to their level of *transparency* (non-transparent, transparent) and *mode of thinking engaged* (automatic mind, reflective mind), following the categories proposed by [119] (see Figure 5.1). The distinction between automatic and reflective nudging approaches is closely related to the *Elaboration Likelihood Model* by Petty and Cacioppo [262]. The Elaboration Likelihood Model is a theoretical framework that distinguishes between the *peripheral* and the *central* route of processing persuasive interventions such as nudges. Automatic nudging, which operates through the peripheral route of processing aims at *influencing behavior* by relying on simple, non-argumentative cues to evoke intuitive and unconscious reactions. Reflective nudging, which operates through the central route of persuasion, aims at *prompting reflective choice* by engaging the critical thinking skills of the recipient to evaluate the arguments presented in a message.

The use of automatic nudges has received criticism for being paternalistic, harming user autonomy, decreasing user experience, hindering learning, and resulting in habituation effects [49, 119, 157]. Yet, purely reflective nudging approaches may not be suitable either in the context of bias mitigation. Processing reflective nudges could further increase cognitive demand and, thus, the susceptibility to cognitive biases.

Prior research on confirmation bias mitigation during web interactions with information items investigated interventions with different objectives: *facilitating information processing*, e.g., with data visualization [201] or argument summaries [300]; *increasing exposure to selected items*, e.g., with preference-inconsistent recommendations [318] or alternative query suggestions [270]; or *raising visibility of behavior*, e.g., with feedback on the political leaning of a user's reading behavior [238].

To mitigate confirmation bias during search result selection, interventions that aim at *decreasing exposure to selected items*, namely attitude-confirming search results, may also be effective. While such interventions have not yet been investigated for confirmation bias mitigation during web search, they have been researched in a different context – to prevent engagement with mis- and disinformation. A particularly successful approach that has been applied across different social networking platforms consists of *warning labels* to flag items that may contain misinformation and *obfuscations* to decrease the ease of access to these items by default [55, 157, 221]. Categorizing these interventions according to the taxonomy by Caraban et al. [49], they combine reflective and automatic nudging elements: they *prompt reflective choice* by confronting users with the risk of engaging with a given item through the warning label and *influence behavior* by decreasing the ease of access to the item through default obfuscations that can be removed with additional effort. Similar interventions that decrease exposure to attitude-confirming items could mitigate confirmation bias during search result selection.

5

5.2.3. Cognitive Reflection and Engagement with Information

Search behavior, susceptibility to cognitive biases, and reaction to nudging approaches are affected by various context-dependent user states and relatively stable user traits. A relatively stable user trait in the context of engagement with information is a user's *cognitive reflection style*. The concept is closely related to the *need for cognition*, an individual's tendency to organize their experience meaningfully [46, 90]. An individual's cognitive reflection style can be captured with the *Cognitive Reflection Test* **(CRT)** [90]. People with a high CRT score are considered to rely more on analytic thinking, thus enjoying challenging mental activities. People with a low CRT score, on the other hand, are considered to rely more on intuitive thinking, thus enjoying effortless information processing [46, 62, 90].

This general tendency of relying on either more analytic or intuitive thinking affects different aspects of engaging with information [47, 261, 348]. Searchers with an analytic cognitive style were observed to invest more cognitive effort in information search [362]. Compared to more intuitive thinkers, analytic individuals were further found to more effectively overcome uncertainties, critically assess their arguments, and monitor their thinking during learning tasks in an online environment [322]. Coutinho [63] found that a more analytic cognitive style is positively correlated with higher metacognitive skills, hence with increased thinking about thinking, a more accurate self-assessment, and increased awareness of one's behavior.

Users' cognitive reflection style was observed to impact whether and how users engage with false information and information that they perceive to be untrustworthy [235, 261, 348]. Tsfati and Capella [348] observed that more analytic people are more likely than intuitive people to engage with information from sources they do not perceive as trustworthy. The authors reason that analytic people do so because they want to make sense of the world and learn about different viewpoints while intuitive people tend to avoid exposure to mistrusted sources. Pennycook and Rand [261] found that analytic users more accurately detect fake news than intuitive users, even if the false information aligns with their ideology. Mosleh et al. [235] observed that intuitive users are generally more gullible (i.e., more likely to share money-making scams and get-rich schemes). They further observed cognitive echo chambers, emerging clusters of accounts of either analytic or intuitive social media users.

Whether people are generally more intuitive or analytic thinkers is a contributing factor to their susceptibility to peripheral (i.e., automatic nudging elements) or central (i.e., reflective nudging elements) cues of persuasion [47]. In the context of nudging, intuitive thinkers might thus be more inclined to follow automatic nudging and choose the path of lowest effort which leads to an unconscious change in their behavior. Analytic thinkers, on the other hand, might be more inclined to follow reflective nudging elements and actively decide to change their behavior.

5.3. Warning Label and Obfuscation Study

With the work presented in this chapter we aim to understand the benefits and risks of an intervention to support unbiased search on debated topics. Therefore, with our first preregistered user study[1], we tested the following hypothesis:[2,3]

H1: *Search engine users are less likely to click on attitude-confirming search results when some search results on the search engine result page (SERP) are displayed with a warning label with obfuscation.*

We conducted a between-subjects user study to test this hypothesis. We manipulated the **search result display** (*targeted warning label with obfuscation, random warning label with obfuscation, regular*) and evaluated participants' **clicks on attitude-confirming search results**. To gain a more comprehensive understanding

[1] The preregistration of this study can be found in our repository: https://osf.io/32wym/?view_only=19cf6003ec1b45c29dbd537058d14b4f.

[2] Next to H1, we tested additional hypotheses on the task design and behavioral patterns across tasks in this user study. The results are not relevant to the focus of this chapter but are presented in Chapter 4.

[3] We reformulated some research questions and hypotheses to ensure consistency in wording across both studies. In terms of content, they remain the same as in the preregistrations.

of the potential benefits and risks of this intervention on search behavior and searchers and uncover potential variations among individuals, we investigated trends in supplementary exploratory data that we collected with this user study. This exploratory data comprises participants' cognitive reflection style, their engagement with the warning label and obfuscated search results (**clicks on show-button**, **clicks on search results with warning labels**), as well as participants' reflection after the interaction (**attitude change**, **accuracy bias estimation**). Note that, throughout the chapter, all analyses labeled as *exploratory* were not preregistered.

5.3.1. Method

Experimental Setup

All related material, including the pre- and post-search questionnaires, can be found at the link in Footnote 1.

Topics and Search Results. The dataset contains search results for the following four debated topics: **(1)** Is Drinking Milk Healthy for Humans? **(2)** Is Homework Beneficial? **(3)** Should People Become Vegetarian? **(4)** Should Students Have to Wear School Uniforms? For each of these, viewpoint and relevance annotations were collected for 50 search results. Out of this dataset of 50 search results per topic, 12 randomly selected search results with overall balanced viewpoints (two *strongly supporting*, two *supporting*, two *somewhat supporting*, two *somewhat opposing*, two *opposing*, and two *strongly opposing*) on one of the four topics were displayed to the participants.

Warning labels and Obfuscation. In the search result display conditions with intervention, results were obfuscated with a warning label, warning of the risk of confirmation bias and advising the participant to select another item (see (1) in Figure 5.2). The warning label included a link to the *Wikipedia* entry on confirmation bias [378] so that participants could inform themselves. To view the obfuscated search result, participants had to click a button, stating they were aware of the risk of confirmation bias.

Cognitive Reflection Test. We measured participants' cognitive style in the post-interaction questionnaire with the cognitive reflection test **(CRT)** [90]. To avoid an effect of familiarity with the three questions of this widely used test, we reworded the three questions in the following way:

1. A toothbrush and toothpaste cost \$2.50 in total. The toothbrush costs \$2.00 more than the toothpaste. How much does the toothpaste cost? *intuitive: \$0.50, correct: \$0.25*

2. If it takes 10 carpenters 10 hours to make 10 chairs, how many hours would it take 200 carpenters to make 200 chairs? *intuitive: 200 hours, correct: 10 hours*

3. On a pig-farm, cases of a pig virus were found. Every day the number of infected pigs doubles. If it takes 28 days for the virus to infect all pigs on the farm, how many days would it take for the virus to infect half of all pigs on the farm? *intuitive: 14 days, correct: 27 days*

Procedure.
The data was collected via the online survey platform *Qualtrics* [283]. As described in Chapter 4, the user study consisted of the three following steps:

(1) Pre-interaction questionnaire: Participants were given the following scenario: *You had a discussion with a relative or friend on a certain topic. The discussion made you curious about the topic, and to inform yourself further, you are conducting a web search on the topic.* They were asked to state their attitude on the four topics on a seven-point Likert scale ranging from *strongly agree* to *strongly disagree* **(prior attitude)**. Subsequently, they were randomly assigned to one of the topics for which they reported to strongly agree or disagree. If they did not report to strongly agree or disagree on any topic, they were randomly assigned to one of the topics for which they reported to agree or disagree. If participants did not fulfill this requirement (i.e., reported weak attitudes on all topics), they were not able to participate further but received partial payment, proportional to the time invested in the task. For the assigned topic, they were asked to state their knowledge on a seven-point Likert scale ranging from *non-existent* to *excellent* **(self-reported prior knowledge)**.

(2) Interaction with the search results: Participants were randomly assigned to one of the three search result display conditions (*targeted warning label with obfuscation, random warning label with obfuscation, regular*) **(search result display)**. Moreover, they were assigned to one out of two task conditions, in which we asked participants to explore the search results by clicking on search results and retrieving the linked documents and mark search results that they considered to be particularly relevant and informative either simultaneously, or in two subsequent steps (for details see [298]). With this chapter, however, we focus exclusively on searchers' exploration (i.e., clicking) behavior. Since we did not find differences in clicking interactions between both task conditions, these conditions are combined into a single group for all subsequent analyses.

For the search task, participants were exposed to 12 viewpoint-balanced search results on their assigned topic. Of those, four search results were initially displayed with a warning label with obfuscation in the targeted and random warning label with obfuscation conditions. To reveal the obfuscated search results, participants could click on a button, from here on referred to as *show-button* **(clicks on show-button)**. From the interaction logs, we calculated the proportion of participants' **clicks on attitude-confirming search results**. For participants in the targeted and random warning label with obfuscation conditions, we calculated the proportion of **clicks on search results with warning labels**. We did not include a time limit in either direction to enable natural search behavior (as far as this is possible in a controlled experimental setting). However, data of participants who did not click on any search result and/or who spent less than one minute exploring the SERP was excluded before data analysis.[4]

(3) Post-interaction questionnaire: Participants were asked to state their attitude again **(attitude change)**. Further, they were asked to reflect and report on their

[4] In a pre-test, we observed that participants who spent less than a minute engaged notably less with the search page. We thus applied the one-minute cut-off to filter out low-quality data from crowdworkers who satisficed by investing minimal effort in the task [171].

search result exploration on a 7-point Likert scale ranging from *all search results I clicked on opposed my prior attitude* to *all search results I clicked on supported my prior attitude* **(accuracy bias estimation)**. To conclude the task, participants were asked to answer the three questions of the CRT **(cognitive reflection)**. To control for data quality, four attention checks were integrated into the task, in which we instructed participants on which response to provide.

Variables

- Independent Variable: **Search result display** (categorical). Participants were randomly assigned to one of three display conditions (see warning label and obfuscation study in Figure 5.3): (1) targeted warning label with obfuscation of extreme attitude-confirming search results, (2) random warning label with obfuscation of four randomly selected search results, and (3) regular (no intervention).

- Dependent Variable: **Clicks on attitude-confirming search results** (continuous). The proportion of attitude-confirming results among the search results participants clicked on during search results exploration.

5

- Exploratory Variables:

  - **Clicks on search results with warning labels** (continuous). For targeted and random warning label with obfuscation conditions: Proportion of obfuscated results among the search results participants clicked on during search results exploration.

  - **Cognitive reflection** (categorical). Participants' cognitive reflection style was measured with an adapted version of the Cognitive Reflection Task (see 5.3.1) in the post-interaction questionnaire. Participants with zero or one correct response were categorized as **intuitive**, and participants with two or three correct responses were categorized as **analytic**.

  - **Clicks on show-button** (discrete). Number of clicks on unique *show-buttons* (up to 4) to reveal an obfuscated search result (only in conditions with obfuscation).

  - **Attitude change** (discrete). Difference between attitude reported on a seven-point Likert scale, ranging from *strongly disagree* (-3) to *strongly agree* (3) in the pre-interaction questionnaire and the post-interaction questionnaire. Attitude difference is encoded in a way that negative values signify a change in attitude towards the opposing direction, whereas positive values indicate a reinforcement of the attitude in the supportive direction. Since we only recruited participants with moderate and strong prior attitudes (-3, -2, 2, 3), the values of attitude change can range from -6 (change from +3 to -3, or -3 to +3) to 1 (change from +2 to +3, or -2 to -3).

  - **Accuracy bias estimation** (continuous). Difference between a) *observed bias* (as the proportion of attitude-confirming clicks) and b) *perceived bias* (reported in the post-interaction questionnaire and re-coded into values from 0 to 1). Values range from -1 to 1, with positive values indicating an overestimation and negative values and underestimation of bias.

- **Self-reported prior knowledge** (discrete). Reported on a seven-point Likert scale ranging from *non-existent* to *excellent* as a response to how they would describe their knowledge on the topic they were assigned to.
- **Usability and Usefulness** (continuous). Mean of responses on a seven-point Likert scale to the modules *usefulness, usability* (six items) from the *meCUE 2.0*[5] questionnaire.

To describe the sample of study participants, we further asked them to report their age and gender.

5.3.2. Results

Description of the Sample.

An a priori power analysis for a between-subjects ANOVA (with $f = 0.25$, $\alpha = \frac{0.05}{4} = 0.0125$ (due to initially testing four different hypotheses, see Footnote 2), and $(1-\beta) = 0.8$) determined a required sample size of 282 participants. Participants were required to be at least 18 years old and to speak English fluently. They were allowed to participate only once and were paid £1.75 for their participation (*mean* = £7.21/h). To achieve the required sample size, we employed a staged recruitment approach, sequentially recruiting participants and monitoring the number of participants that fulfill the inclusion criteria detailed below. For that, we recruited a total of 510 participants via the online participant recruitment platform *Prolific* [277]. From these 510 participants, 228 were excluded from data analysis for failing the following preregistered inclusion criteria: they did not report having a strong attitude on any of the topics *(41)*, failed at one or more of four attention checks *(50)*, spent less than 60 seconds on the SERP (*80*), or did not click on any search results *(57)*. We paid all participants regardless of whether we excluded their data from the analysis.

Our final data-set consisted thus of 282 participants, of which 51% reported to be male, 49% female, <1% non-binary/other. Concerning the age of the participants, 49.6% reported to be between 18 and 25, 27.3% between 26 and 35, 12.1% between 36 and 45, 7.1% between 46 and 55, 3.5% between 56 and 65, and 0.4% more than 65 years old.

The task in each display condition was completed by 80 to 102 participants and 58 to 85 participants saw search results of the different topics (see Table 5.1). The mean time spent exploring the SERP was 4min 45sec ($SE = 15.6sec$), ranging from a minimum of 1 min to a maximum of 26 min, with no evidence for differences between search result display conditions ($F(2,279) = 0.34, p = .71, f = 0.05$). The mean number of clicks on search results was 3.26 ($SE = 0.13$), approximately 25% of the 12 displayed search results, with no evidence for differences between search result display conditions ($F(2,279) = 0.88, p = .42, f = 0.08$).

[5] meCUE usability scale: http://mecue.de/english/home.html

Table 5.1: Distribution across conditions in warning label and obfuscation study: Number of participants per search result display conditions and topic (1: Is Drinking Milk Healthy for Humans?; 2: Is Homework Beneficial?; 3: Should People Become Vegetarian?; 4: Should Students Have to Wear School Uniforms?).

	Topic 1	Topic 2	Topic 3	Topic 4	All
targeted w + obf	20	32	19	31	**102**
random w + obf	23	27	19	31	**100**
regular	15	22	20	23	**80**
All	**58**	**81**	**58**	**85**	**282**

HYPOTHESIS TESTING: EFFECT OF SEARCH RESULT DISPLAY ON CLICKS ON ATTITUDE-CONFIRMING SEARCH RESULTS

Although the distribution of attitude-confirming clicks did not exhibit normality, it is worth noting that ANOVAs have shown robustness in studies involving large sample sizes, even in cases where normality assumptions are not met [35, 379]. Considering this, we opted to employ ANOVAs for the statistical assessment of variations in participants' click behavior. The results of the ANOVA show evidence for a moderate effect of **search result display** on clicks on attitude-confirming search results ($F(2,279) = 17.14, p < .001, f = 0.35$).[6] A pairwise post-hoc Tukey's test shows that the proportion of clicks on attitude-confirming search results was significantly lower for participants who were exposed to **targeted warning labels with obfuscations** ($mean = 0.34, SE = 0.03$) compared to those who saw **random warning labels with obfuscations** ($mean = 0.55, SE = 0.03;\ p < .001$), and those who saw **regular** search results ($mean = 0.58, SE = 0.03;\ p < .001$; see Figure 5.4). However, there was no evidence for a difference in the clicking behavior between **random warning labels with obfuscations** and **regular** search result display.

EXPLORATORY OBSERVATIONS.

We inspected the exploratory data to derive new hypotheses by visually investigating plots of means and standard errors, as well as boxplots of the (exploratory) dependent variables **clicks on search results with warning labels**, **clicks on show-button**, **attitude change**, and **accuracy bias estimation** for the (exploratory) independent variables **search result display** and **cognitive reflection**. We observed that participants who, according to the CRT, are more analytic thinkers were more likely to engage with search results with warning labels and to click on the show-button (see Figures 5.5 and 5.6). Further, participants' attitude change seemed to be influenced by the display condition and their cognitive reflection style (see Figure 5.7). We also noted that participants who were exposed to targeted warning labels with obfuscations tended to overestimate their confirmation bias. Analytic participants more accurately estimated their bias than intuitive participants (see Figure 5.8).

We further explored means and standard errors of clicks on attitude-confirming search results across different degrees of self-reported **prior knowledge**, yet no

[6] We validated the ANOVA results by additionally applying a Kruskal-Wallis test which likewise yielded a moderate effect ($H(2) = 33.87, p < .001, \eta^2 = 0.11$)

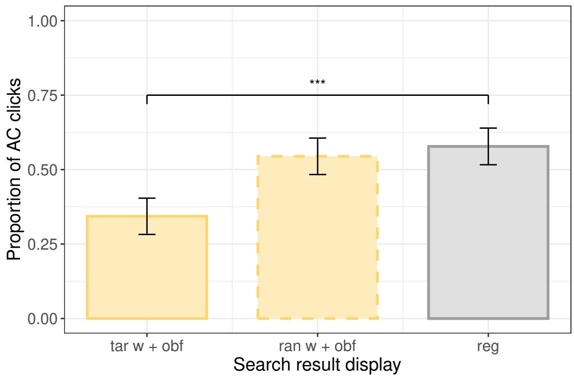


Figure 5.4: Study 1: **Clicks on attitude-confirming search results**. Mean proportion of participants' attitude-confirming clicks per search result display condition (targeted warning label with obfuscation, random warning label with obfuscation, regular) with 95% confidence intervals. A proportion of one implies that all clicks were on attitude-confirming search results.

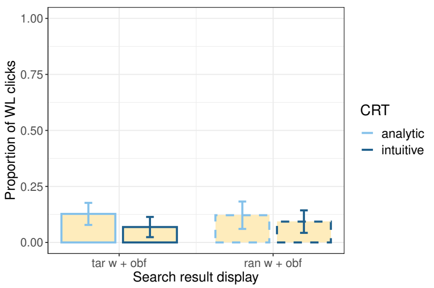


Figure 5.5: Study 1 (exploratory): **Clicks on search results with warning labels**. Mean proportion of clicks on search results that were displayed with a warning label per search result display condition (targeted warning label with obfuscation, random warning label with obfuscation) and cognitive reflection style (analytic, intuitive) with 95% confidence intervals.

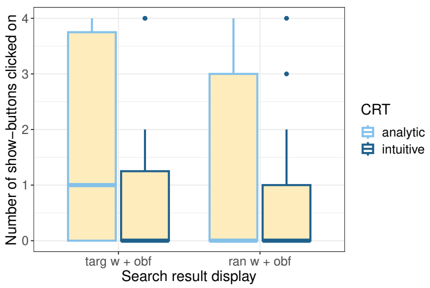


Figure 5.6: Study 1 (exploratory): **Engagement with warning labels**. Boxplots with medians and quartiles, illustrating the distribution of the number of show-buttons that each participant clicked on (up to four) per search result display condition (targeted warning label with obfuscation, random warning label with obfuscation) and cognitive reflection style (analytic, intuitive).

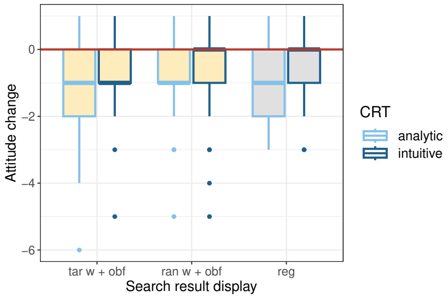


Figure 5.7: Study 1 (exploratory): **Attitude change**. Boxplots with medians and quartiles, illustrating the distribution of participants' difference between pre- and post-interaction attitude per search result display condition (targeted warning label with obfuscation, random warning label with obfuscation, regular) and cognitive reflection style (analytic, intuitive). Negative values indicate a weakening of the initial attitude.

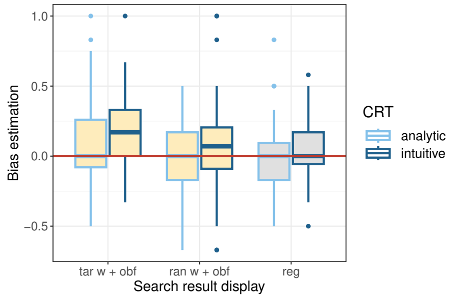


Figure 5.8: Study 1 (exploratory): **Accuracy of bias estimation**. Boxplots with medians and quartiles, illustrating the distribution of participants' difference between observed bias and perceived bias per search result display condition (targeted warning label with obfuscation, random warning label with obfuscation, regular) and cognitive reflection style (analytic, intuitive). Positive values indicate an overestimation of bias (i.e., perceived bias is higher than observed bias in behavior).

differences emerged. Finally, we investigated whether participants in distinct **search result display** conditions exhibited different levels of usefulness and usability. The inspection of means and standard errors revealed no discernible differences between the three conditions (see Table 5.2).

Table 5.2: Study 1 (exploratory): **Usability and Usefulness**. Mean usability and usefulness scores with standard error per search result display condition (targeted warning label with obfuscation, random warning label with obfuscation, regular)

	Usability		**Usefulness**	
	mean	*SE*	*mean*	*SE*
Targeted w + obf	6.06	0.09	5.47	0.1
Random w + obf	6	0.1	5.52	0.11
Regular	6.24	0.11	5.59	0.11

5.3.3. Reflections and Follow-up Hypotheses

We found that targeted obfuscations with warning labels decreased the likelihood of clicking on attitude-confirming search results. However, it is unclear whether the intervention prompted *reflective* choice, and participants read the warning label and clicked on the show-button to reveal the search result but, now aware of confirmation bias, actively decided to interact less with attitude confirming search results; or the intervention *automatically* influenced behavior, and participants engaged less with obfuscated items because interaction with those required additional effort.

Our exploratory findings indicate that both targeted and random warning labels decrease engagement with search results with warning labels and that intuitive searchers are less likely to engage with the warning label by clicking on the show-button than analytic searchers. This could imply that, in line with the Elaboration Likelihood Model [262], for more intuitive users, decreased engagement might be caused primarily by the obfuscation. Yet, if intuitive users do not engage with the intervention and ignore the warning label, the intervention might effectively not be transparent and manipulate instead of influence user behavior (see Figure 5.1).

To understand how different searchers are impacted by the *reflective* and *automatic* elements of the intervention, we need to investigate the effects of warning labels and obfuscations separately (**warning labels with and without obfuscations**). Based on our exploratory insights, we suggest the following primary hypotheses[3] for this follow-up study:

- **H2a:** Search engine users are less likely to click on search results that are displayed with a warning label with obfuscation than search results that are displayed with a warning label without obfuscation.

- **H2b:** Intuitive search engine users are less likely to click on a button to reveal an obfuscated search result than analytic users.

- **H2c:** The difference in clicks on search results that are displayed with a warning label without obfuscation compared to those with obfuscation is moderated by users' cognitive reflection style.

- **H2d:** Clicks on search results that are displayed with a warning label with obfuscation will be reduced, while clicks on search results with a warning label without obfuscation will only be reduced when they are applied to attitude-confirming search results (targeted) but not when they are applied incorrectly, to random search results.

- **H2e:** The moderating effect of targeting on the effect of warning style on users' clicks on search results with warning labels is moderated by users' cognitive reflection style.

Further, based on our exploratory observations on attitude change and accuracy of bias estimation, we suggest the following secondary hypotheses:[3]

- **H3a:** Attitude change is greater in conditions with targeted warning labels than in conditions with random warning labels and no warning labels.

- **H3b:** The effect of the search result display condition on attitude change is moderated by participants' cognitive reflection style.

- **H4a:** Users who see search results with targeted warning labels overestimate the confirmation bias in their clicking behavior to a greater extent than users who see search results with random or no warning labels.

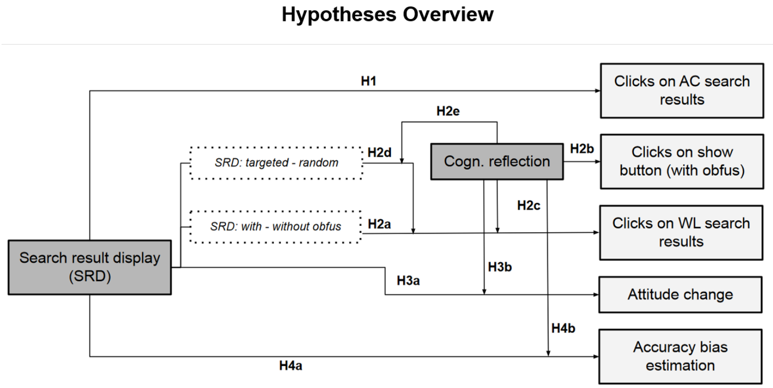


Figure 5.9: Study 2: Automatic vs. reflective study. Overview of hypotheses with independent (dark grey) and dependent (light grey) variables.

5

- **H4b:** Analytic participants make more accurate estimations of the bias in their behavior while intuitive participants tend to overestimate the bias in their behavior.

5.4. FOLLOW-UP: AUTOMATIC VS. REFLECTIVE STUDY

We conducted a follow-up study, the **automatic vs. reflective study**, with the primary goal to better understand the effect of warning labels and obfuscations on different users' search behavior. Specifically, we investigated whether the observed effect was caused by the obfuscation (automatic) or the warning label (reflective) (H2a, H2d). With this follow-up study, we also tested whether we could replicate the findings we made in the **warning label and obfuscation study** for different search results, but the same topics (H1). To better understand the impact of the interventions on the searcher, we further tested whether the search result display has effects on their attitude change (H3a) and awareness of bias (H4a). Finally, we investigated the potential (moderating) effects of participants' tendency to be more intuitive or analytic thinkers, according to their CRT scores, on their engagement with the intervention (H2b), engagement with search results with warning labels (H2c, H2e), attitude change (H3b) and accuracy of bias estimation (H4b) (see Section 5.3.3 and Figure 5.9).

5.4.1. METHOD

The method we used for the second, preregistered[7], between-subjects user study was essentially identical to the method we used for the first user study. We made the following minor changes to permit testing the follow-up hypotheses (H2-H4, see Section 5.3.3):

[7]The preregistration of the second user study can be found in our repository: https://osf.io/p3ykv/?view_only=93f2eebbd55445aea3604ae751127892.

- **Search result display**: To allow us to understand the distinct impact of the *automatic* (obfuscation), and the **reflective** (warning label) nudging element of the intervention, we introduced two additional **search result display** conditions: targeted and random warning label without obfuscation (see (2) in Figure 5.2). This resulted in the following five display conditions (see Figure 5.3):

  1. **targeted** warning label **with obfuscation** of moderate and extreme attitude confirming search results
  2. **targeted** warning label **without obfuscation** of moderate and extreme attitude confirming search results
  3. **random** warning label **with obfuscation** of four randomly selected search results
  4. **random** warning label **without obfuscation** of four randomly selected search results
  5. regular (no intervention)

- **Experimental Setup**: To test the reproducibility of the findings in the **warning label and obfuscation study** for different search results, we randomly sampled new search results (12 per topic, two *strongly supporting*, two *supporting*, two *somewhat supporting*, two *somewhat opposing*, two *opposing*, two *strongly opposing*) for the same topics from the set of viewpoint annotated search results which we collected for the **warning label and obfuscation study**.

  Since concerns about the validity of the CRT have been raised [116, 340], we included the exploratory variable of participants' *need for cognition*, a measure that captures users' motivation to engage in effortful thinking, to support potential findings on moderating effects of cognitive reflection. We captured participants' need for cognition with a self-report with a 4-item subset of the need for cognition questionnaire by Cacioppo et al. [46]. These four items include the same subset as used in Buçinca et al. [42]: *I would prefer complex to simple problems; I like to have the responsibility of handling a situation that requires a lot of thinking; Thinking is not my idea of fun; I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.*

- **Variables**: Exploratory variables in the **warning label and obfuscation study** were turned into independent and dependent variables in the **automatic vs. reflective study**. In the **automatic vs. reflective study**, we thus manipulated and measured the following variables:

  – **Independent Variables**: Search result display, cognitive reflection

  – **Dependent Variables**: Clicks on attitude-confirming search results (attitude-confirming), clicks on search results with warning labels, clicks on show-button, attitude change, accuracy bias estimation

  - **Exploratory Variables**: Need for cognition, prior knowledge, usability and usefulness

- **Procedure**: The procedure of data collection remained essentially the same as described in Section 5.3.1 for the **warning label and obfuscation study**. The four questions to capture the need for cognition were added to the post-interaction questionnaire. We slightly increased the reward for participation to 1.80£ (mean = 7.89£/h) to adhere to the updated Prolific recommendations for fair pay. Further, we launched the data collection in multiple batches at different times of the day and night, to increase the likelihood of a sample with high diversity in geographical locations.

- **Attention checks**: To adhere to Prolific guidelines, we included an additional attention check, leading to a total of five, and adapted the exclusion criterion to failing two or more (instead of one or more out of four) attention checks.

5

5.4.2. Results

Description of the Sample.

An a-priori power analysis for between-subjects ANOVAs, assuming moderate effects ($f = 0.25$, $\alpha = \frac{0.05}{10} = 0.005$ (due to testing 10 hypotheses), $(1-\beta) = 0.8$, up to 10 groups) determined a required sample size of 307 participants. As for the **warning label and obfuscation study**, we employed a staged recruitment approach in which we recruited an overall of 481 participants. Of these, 174 were excluded because they did not fulfill the inclusion criteria: they did not report having a strong attitude on any of the topics *(31)*, failed at two or more of five attention checks *(2)*, spent less than 60 seconds on the SERP*(88)*, or did not click on any search results (*53*). Of the 307 included participants, 52% reported to be male, 46% female, 2% non-binary/other, and <1% preferred not to share their gender. Further, 40.7% reported to be between 18 and 25, 37.1% between 26 and 35, 12.7% between 36 and 45, 6.8% between 46 and 55, 1.6% between 56 and 65, and 1% more than 65 years old.

53 to 67 participants completed the task in each of the five search result display conditions and 46 to 93 participants saw search results for each of the four topics (see Table 5.3). Results of the CRT categorized 167 participants as *analytic* and 140 participants as *intuitive*. The mean time spent exploring the SERP page was 4 min 19 sec ($SE = 10.2sec$), ranging from a minimum of 1 min to a maximum of 19 min, with no evidence for differences between search result display conditions ($F(4,302) = 0.57, p = .69, f = 0.09$) and cognitive reflection categories ($F(4,302) = 2.18, p = .14, f = 0.08$). The mean number of clicks on search results was 2.8 ($SE = 0.09$), with no evidence for differences between search result display conditions ($F(4,302) = 1.24, p = .29, f = 0.13$), but a difference between cognitive reflection categories ($F(4,302) = 18.09, p < .001, f = 0.24$) with more clicks by analytic ($mean = 3.16, SE = 0.14$) than intuitive ($mean = 2.38, SE = 0.12$) participants.

Table 5.3: Distribution across conditions in automatic vs. reflective study: Number of participants per search result display conditions and topic (1: Is Drinking Milk Healthy for Humans?; 2: Is Homework Beneficial?; 3: Should People Become Vegetarian?; 4: Should Students Have to Wear School Uniforms?).

	Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**All**
targeted w + obf	18	19	7	18	**62**
targeted w	18	16	14	19	**67**
random w + obf	11	25	11	15	**62**
random w	11	13	9	20	**53**
regular	25	20	5	13	**63**
All	**83**	**93**	**46**	**85**	**307**

HYPOTHESIS TESTING.

We conducted five ANOVAs to test the ten hypotheses and set the significance threshold at $\alpha = \frac{0.05}{10} = 0.005$, aiming at a type 1 error probability of $\alpha = 0.05$ and applying Bonferroni correction to correct for multiple testing.

H1: Main effect of search result display on attitude-confirming clicks (Replication). We could replicate the findings made in the **warning label and obfuscation study** by finding more evidence for a moderate effect of the **search result display** on **clicks on attitude-confirming search results** ($F(4,302) = 6.67, p < .001, f = 0.30$). A pairwise posthoc Tukey's test shows that the proportion of clicks on attitude-confirming search results was significantly lower for participants who were exposed to targeted warning labels with obfuscations ($mean = 0.34, SE = 0.03$) than those who were exposed to a regular search page ($mean = 0.53, SE = 0.04; p = .004$; see Figure 5.10). In comparison to the regular search page, participants exposed to targeted warning labels without obfuscations likewise exhibited a lower mean proportion of clicks on attitude-confirming search results ($mean = 0.41, SE = 0.03$). As in the **warning label and obfuscation study**, we did not observe lower proportions of clicks on attitude-confirming search results for participants exposed to random warning labels with obfuscations ($mean = 0.56, SE = 0.05$).

H2a: Main effect of obfuscation on clicks on search results with warning labels. We found evidence for a moderate effect of **obfuscation** on the proportion of clicks on search results that were displayed with a warning label ($F(1,236) = 12.9, p < .001, f = 0.23$). A posthoc Tukey test revealed that in conditions with obfuscations, participants clicked on fewer search results that were displayed with a warning label ($mean = 0.12, SE = 0.02$) than in conditions without obfuscations ($mean = 0.24, SE = 0.03; p < .001$; see Figure 5.11). Thus, H2a was confirmed.

H2b: Main effect of cognitive reflection on clicks of show-button. Descriptive statistics indicated that participants with an analytic as opposed to an intuitive **cognitive reflection** style were more likely to click on the show-button to reveal search results that were initially obfuscated (see Figure 5.12). However, evidence for this relation did not meet the Bonferroni-corrected significance threshold of $\alpha = 0.005$ ($F(1,122) = 6.22, p = .014, f = 0.23$). To gain further insights, we explored (i.e., this analysis was not preregistered) the proportion of participants that did not at all engage with the warning label by clicking on the show-button and observed

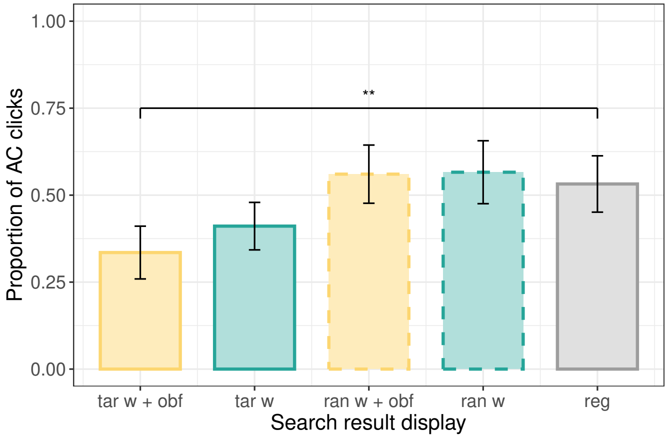


Figure 5.10: Study 2: **Clicks on attitude-confirming search results**. Mean proportion of participants' attitude-confirming clicks per search result display condition (targeted warning label with obfuscation, targeted warning label without obfuscation, random warning label with obfuscation, random warning label without obfuscation, regular) with 95% confidence intervals. A proportion of one implies that all clicks were on attitude-confirming search results.

5

that overall, a high proportion of participants did not even once click on the show-button (56%). This exploratory analysis further revealed that more intuitive (68%) than analytic (47%) participants, and more participants in the random warning label condition (65%) than in the targeted warning label condition (48%) ignored the warning labels (see Table 5.4).

H2c: Interaction effect of cognitive reflection and obfuscation on clicks on search results with warning labels. We did not find evidence for an interaction effect of **cognitive reflection** and **obfuscation** on the proportion of clicks on search results that were displayed with a warning label ($F(1,236) = 0.04, p = .85, f = 0.01$; see Figure 5.11).

H2d: Interaction effect of targeting and obfuscation on clicks on search results with warning labels. Descriptive statistics suggest a disparity of the mean proportion of clicks on search results with warning labels between the conditions with and without obfuscations. This disparity was more pronounced in the random than in the targeted warning labels condition (see Figure 5.11). Yet, the interaction between **targeting** and **obfuscation** did not meet the Bonferroni-corrected significance threshold of $\alpha = 0.005$ ($F(1,236) = 5.41, p = .02, f = 0.15$).

H2e: Interaction effect of cognitive reflection, targeting, and obfuscation on clicks on search results with warning labels. We did not find evidence for an interaction effect of **cognitive reflection**, **targeting**, and **obfuscation** on the proportion of clicks on search results that were displayed with a warning label ($F(1,236) = 0.15, p = .70, f = 0.03$; see Figure 5.11).

H3a: Main effect of search result display on attitude change. We did not find evidence for an effect of **search result display** on participants' **attitude change** ($F(4,297) = 1.55, p = .18, f = 0.14$; see Figure 5.13).

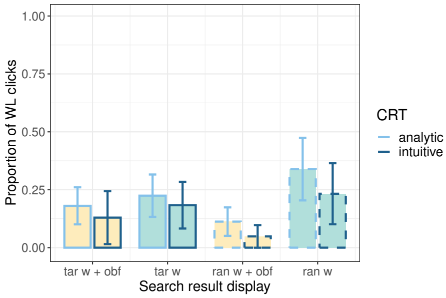


Figure 5.11: Study 2: **Clicks on search results with warning labels**. Mean proportion of clicks on search results that were displayed with a warning label per search result display condition (targeted warning label with obfuscation, targeted warning label without obfuscation, random warning label with obfuscation, random warning label without obfuscation) and cognitive reflection style (analytic, intuitive) with 95% confidence intervals.

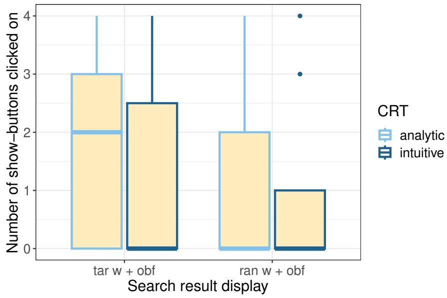


Figure 5.12: Study 2: **Engagement with warning labels** (only for display conditions with obfuscation). Boxplots with medians and quartiles, illustrating the distribution of the number of show-buttons that each participant clicked on (up to four) per search result display condition (targeted warning label with obfuscation, random warning label with obfuscation) and cognitive reflection style (analytic, intuitive).

5

Table 5.4: Study 2 (exploratory): **No engagement with warning labels.** The proportion of participants who did not engage with any warning label by clicking on the show-button per search result display condition (targeted warning label with obfuscation, random warning label with obfuscation) and cognitive reflection style (analytic, intuitive).

	CRT: analytic	**CRT: intuitive**	**All**
targeted w + obf	37%	63%	**48%**
random w + obf	58%	72%	**65%**
All	**47%**	**68%**	

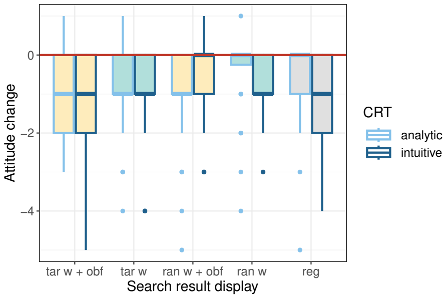


Figure 5.13: Study 2: **Attitude change**. Boxplots with medians and quartiles, illustrating the distribution of participants' difference between pre- and post-interaction attitude per search result display condition (targeted warning label with obfuscation, targeted warning label without obfuscation, random warning label with obfuscation, random warning label without obfuscation, regular) and cognitive reflection style (analytic, intuitive). Negative values indicate a weakening of the initial attitude.

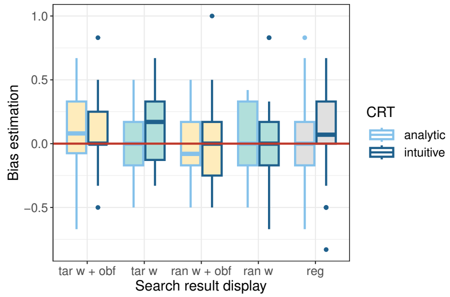


Figure 5.14: Study 2: **Accuracy of bias estimation**. Boxplots with medians and quartiles, illustrating the distribution of participants' difference between observed bias and perceived bias per search result display condition (targeted warning label with obfuscation, targeted warning label without obfuscation, random warning label with obfuscation, random warning label without obfuscation, regular) and cognitive reflection style (analytic, intuitive). Positive values indicate an overestimation of bias (i.e., perceived bias is higher than observed bias in behavior).

H3b: Interaction effect of cognitive reflection and search result display on attitude change. We did not find evidence for an interaction of **cognitive reflection** and **search result display** on **attitude change** did not meet the Bonferroni-corrected significance threshold of $\alpha = 0.005$ ($F(4,297) = 2.72, p = .03, f = 0.19$); see Figure 5.13).

H4a: Main effect of search result display on accuracy of bias estimation. We did not find evidence for an effect of **search result display** on participants' **accuracy of bias estimation** ($F(4,297) = 0.77, p = .55, f = 0.10$; see Figure 5.14).

H4b: Interaction effect of cognitive reflection and search result display on accuracy of bias estimation. We did not find evidence for an interaction effect of **cognitive reflection** and **search result display** on participants' **accuracy of bias estimation** ($F(4,297) = 0.62, p = .64, f = 0.09$; see Figure 5.14).

EXPLORATORY OBSERVATIONS.

To gain deeper insights and support our findings from hypotheses testing, we explored the correlation between CRT and need for cognition, the potential effects of self-reported prior knowledge on engagement behavior and search consequences, and potential differences in usability and usefulness of the different search result display conditions for searchers with an analytic or intuitive cognitive reflection style. We calculated the *Spearman's* correlation coefficient between participants' CRT (behavioral) and need for cognition (questionnaire) score and found a weak positive relationship between the variables ($r = .21$, $p < .001$). Further, we did not observe differences in any of the dependent variables between participants who reported a high compared to a low level of self-reported *prior knowledge*. Lastly, we did not observe any differences in questionnaire-reported usefulness and usability

Table 5.5: Study 2 (exploratory): **Usability and Usefulness**. Mean usability and usefulness scores with standard error per search result display condition (targeted warning label with obfuscation, targeted warning label without obfuscation, random warning label with obfuscation, random warning label without obfuscation, regular) and cognitive reflection style (analytic, intuitive).

	Usability				**Usefulness**			
	CRT: analytic		**CRT: intuitive**		**CRT: analytic**		**CRT: intuitive**	
	mean	*SE*	*mean*	*SE*	*mean*	*SE*	*mean*	*SE*
Targeted w + obf	5.75	0.18	6.14	0.16	5.43	0.17	6.17	0.17
Targeted w	5.99	0.12	6.01	0.19	5.51	0.18	5.96	0.15
Random w + obf	6.14	0.12	6.29	0.1	5.88	0.13	6.09	0.14
Random w	6.04	0.17	5.89	0.17	5.81	0.17	5.74	0.16
Regular	6.21	0.13	6.29	0.09	5.98	0.15	6	0.18

between the five *search result display* conditions. However, there was a tendency of participants who were categorized as *analytic* according to their CRT results to report lower usefulness of the SERP with targeted warning labels with and without obfuscations than participants who were categorized as *intuitive* (see Table 5.5). For the random and regular search result display conditions, no such difference was observed.

Participants who did not click on search results. The high rate of participants who had to be excluded from hypotheses testing because they did not click any search results ($N = 104$) prompted us to investigate possible causes. Our exploration revealed that there were no discernible differences in prior attitude strength or cognitive reflection style between the participants who clicked on search results and those who did not. Furthermore, the results indicate that participants who did not click on any search results were just as likely to change their attitude ($mean = -1.01, SE = 0.12$) as those who did click on one or more search results ($mean = -0.84, SE = 0.06$).

5.5. DISCUSSION

The two pre-registered user studies contribute to the understanding of behavioral interventions to support thorough and unbiased information-seeking strategies that are required for responsible opinion formation on debated topics. Specifically, we focused on mitigating confirmation bias during search result selection by reducing engagement with attitude-confirming search results. Inspired by interventions to reduce engagement with misinformation, we applied warning labels and obfuscations to attitude-confirming search results. We further investigated the risks of the interventions by including conditions in which they were applied incorrectly, to random instead of attitude-confirming search results. To gain more comprehensive insights into the potential effects of the interventions, we did not only investigate participants' search behavior but also their attitude change and awareness of bias. We further investigated the potential moderating effects of participants' cognitive reflection style. The following paragraphs summarise and discuss the findings and observations from both studies. Based on these findings, we discuss implications for designing interventions that aim at supporting thorough and unbiased information-seeking strategies.

5.5.1. Findings and Observations

Warning label and Obfuscation

In the **warning label and obfuscation study**, we found that the intervention effectively reduced engagement. However, it reduced engagement with all search results that it was applied to, even if it was applied incorrectly to search results that were not attitude-confirming. This suggests that the intervention could be misused to manipulate engagement with information for alternative purposes, raising substantial ethical concerns.

The experimental setup did not allow for conclusions on how much of the effect was caused by the warning label (reflective element) versus the obfuscation (automatic element). To investigate the potential effects of both nudging elements separately, we conducted a follow-up study and added a second intervention: We exposed participants to **warning labels without obfuscation** (see (2) in Figure 5.2).

Automatic vs. Reflective

We tested two interventions in the **automatic vs. reflective study**: warning label with obfuscation (reflective and automatic) and warning label without obfuscation (reflective). As before, we tested the interventions on either targeted attitude-confirming or random search results.

Answering RQ1, the mean proportion of clicks on attitude-confirming search results was reduced by targeted warning labels with and without obfuscations. This indicates that the mere warning label, thus the reflective element of the initial intervention, successfully achieves a reduction of clicks on attitude-confirming search results and thus mitigates confirmation bias. Thus, contrary to our concerns, the purely reflective intervention did not exhaust users' processing capacities.

The warning label alone, as opposed to with obfuscations, did not reduce clicks when they were applied incorrectly to random search results. Therefore, it seems that the automatic element is the reason why searchers fail to detect and react to incorrect applications. These findings suggest that obfuscation restricts agency and harms autonomy. This is further supported by the high proportion of participants who seemed to have ignored the warning labels since they did not click on any show-button. While the intervention was designed with the intention to transparently influence behavior and prompt reflective choice, it might effectively manipulate behavior for users who do not engage with it.

These findings are in line with observations that users approach web search on debated topics with the intention to engage with diverse viewpoints [4, 219] but often fail to do so. For instance, Smith and Rieh [325] discuss that users have learned to trust that the resources provided by search engines, especially highly ranked results, are accurate and reliable. The authors reason that this might cause them to exert less cognitive effort in the search process. Yet, for complex search tasks that affect opinion formation, cognitive effort to engage with, compare, and evaluate different viewpoints would be required to form opinions responsibly [226].

Thus, interventions should encourage users to invest more effort into the search process to achieve their intended behavior of engaging with diverse viewpoints.

Cognitive Reflection Style

According to the Elaboration Likelihood Model [262] analytic thinkers might be more likely to follow reflective nudging elements, while intuitive thinkers might be more likely to follow automatic nudging elements. Thus, we investigated the potential moderating effects of participants' cognitive reflection style on their engagement behavior.

In the **automatic vs. reflective study**, we did not find evidence for significant differences in engagement with the search results and interventions between users who, according to their CRT scores, are more analytic or intuitive thinkers. However, we did observe that, in line with the Elaboration Likelihood Model [262], the proportion of participants who did not at all engage with the warning labels is higher for intuitive (68%) than for analytic (47%) thinkers.

We attribute the lack of evidence for a moderating effect of cognitive reflection style on clicks on the show-button on a combination of high noise in our data and strictly Bonferroni-corrected significance thresholds. The noise might have been caused by other user and context factors, such as their prior knowledge, situational and motivational influences (e.g., metacognitive states or traits), and ranking effects. In light of these considerations, a conclusive answer to RQ2 is still pending. Future research should thus continue to investigate the potential effects of users' cognitive reflection style and other user traits, states, and context factors that might moderate the effects of automatic and reflective elements of a nudge.

Attitude Change and Awareness of Bias

To gain more comprehensive insights into the potential effects of the intervention, we compared users' attitude change and awareness of bias between the different search result display conditions and cognitive reflection styles. Answering RQ3 and RQ4, we neither found evidence for differences between search result display conditions in participants' attitude change and awareness of bias nor for moderating effects of participants' cognitive reflection style. For both variables, we observed high levels of noise that might be caused by user differences beyond their cognitive reflection style.

In terms of responsible opinion formation, participants' prior knowledge of the topic should have a great impact on their attitude change. Users who have well-rounded prior knowledge should be less likely to change their attitude since it was already formed responsibly. Thus, it is unclear whether and what direction of attitude change would indicate responsible opinion formation.

Regarding awareness of bias, relatively stable traits and context-dependent states of users' metacognition (i.e., thinking about one's thinking) would likely have an impact and might have caused some of the observed noise. Of particular interest for responsible opinion formation and the risk of confirmation bias is users' *intellectual humility*, their ability to recognize the fallibility of their beliefs, and the limits of their knowledge [66, 267, 295]. Compared to people with low intellectual humility, those with high intellectual humility were observed to invest more effort in information-seeking, spend more time engaging with attitude-opposing arguments [182, 268], and more accurately recognize the strength of different arguments, regardless of their stance [195]. Thus, high intellectual humility appears to reduce the likelihood of behavioral patterns that are common for confirmation bias [295]. The effect of metacognitive traits and states on search behavior and responsible opinion formation should be investigated in future research.

5.5.2. Implications

The observations and considerations discussed in the previous sections illustrate the complexity of researching and supporting web search for responsible opinion formation. The intervention of warning labels with obfuscations was inspired by approaches to combat misinformation. While we investigated this intervention because some objectives of combating misinformation overlap with those of mitigating confirmation bias during search, the research process and findings made us aware of a fundamental difference between them. Misinformation is a user-external threat and user behavior that is desired by system designers is fairly clearly defined (reduced/no engagement with items that contain misinformation). This is not the case for cognitive biases that impact search for opinion formation, which are user-internal and, depending on the context, serve a function [102].

As interventions to combat misinformation, the interventions we tested primarily aimed at reducing engagement with selected information items. To mitigate confirmation bias during search result selection, we aimed to reduce engagement with attitude-confirming search results. However, it is unclear what proportion of engagement with different viewpoints is desirable to support responsible opinion formation. When wanting to support users in gaining a well-rounded knowledge, the desirable proportion likely depends on users' prior knowledge of the arguments for the different viewpoints. This illustrates that what constitutes *beneficial behavior* for responsible opinion formation during search on debated topics is non-trivial to define due to complex context and user dependencies.

Aiming for interventions that decide which information should be engaged with on the users' behalf imposes an immense level of responsibility on authorities who design them and decide on the application criteria [30]. Such interventions harm user autonomy and provide the means for abuse with intentions of stirring user behavior with (malicious) interests that do not align with the user's own interests. In preparation for our studies, we justified these risks of applying an automatic nudging element with the aim of reducing users' cognitive processing load. In fact, however, this was not necessary since users chose to engage less with attitude-confirming search results when prompted to do so by a warning label without obfuscation. Thus, we may be underestimating users' abilities to actively choose unbiased behavior. Therefore, the risks of applying automatic nudging elements to support thorough information-seeking strategies are likely unwarranted. This potentially applies to other nudging scenarios in which the desired behavior is not clearly defined but depends on various (unknown) context and user factors.

Design Guidelines for Interventions. Given the complexity and potential far-reaching impact of search for opinion formation, we argue that interventions to support thorough and unbiased search should strictly emphasize user agency and autonomy. As a practical consequence, nudging interventions should prioritize reflective and transparent elements.

As an alternative to nudging interventions that steer user behavior directly, encouraging thorough information-seeking strategies could also be achieved by educating and empowering users to actively choose to change their behavior [295]. This can be done with boosting interventions that attempt to teach users to become resistant to various pitfalls of web interactions and remain effective for some time after being exposed to the intervention [132, 210]. Such approaches would improve user autonomy, minimize the risk of abuse and errors, and tackle the factors that impede search for responsible opinion formation more comprehensively and sustainably [100, 132, 178, 210, 295]. Next to boosting, thorough information-seeking strategies that entail exploring, comparing, or evaluating different resources for sense-making and learning could be supported by other means of designing the search environment (e.g., adding metadata, such as stance labels) [324, 325].

Whether nudging, boosting, or other approaches, interventions that aim at supporting search for responsible opinion formation should be designed to increase transparency to and choice for the user [398]. This claim aligns with the EU's ethics guideline for trustworthy AI, which places human autonomy and agency at its core and states that AI systems (e.g., search engines) should support humans to make informed decisions by augmenting and complementing human cognitive skills instead of manipulating or herding them [85].

5.5.3. Limitations and Future Work

We acknowledge some limitations, mainly resulting from the controlled setting of this user study. We chose the controlled setting to be able to clearly distinguish the effects of the interventions from other factors that might affect search behavior. For that, we constructed an artificial scenario with one specific search task. Further, we presented one specific set of pre-selected topics and viewpoint-labeled search results on a single SERP. While our objective was to closely assimilate real-world search settings, this controlled experimental setup did not allow participants to issue multiple queries or have access to great amounts of resources over an extended time period. Further, while assigning participants to a topic for which they reported a strong attitude, we did not capture whether they were interested in learning about it. Future research should investigate whether the effects we observed will also be observed in less controlled search settings, how they evolve when users are exposed to the interventions for multiple search sessions, and whether the effects of the intervention are different for searchers who report weak prior attitudes on the topics.

We further attempted to ensure that ranking effects (i.e., position bias that causes more engagement with high-ranked items [110, 151]) would not distort the effects of the search result display by fully randomizing the ranking. Yet, given these known strong effects of search result ranking on user engagement, this design decision might have added noise to our data that prevented us from finding significant evidence for some of our hypotheses. Future work should thus investigate the interplay of interventions with ranking effects during search on debated topics.

Our representation of prior knowledge was limited. We did anticipate that prior knowledge could affect users' search behavior [93, 334] and attitude change, especially for users with strong opinions on debated topics. We thus captured users' self-reported prior knowledge. However, we did not find any effects of self-reported prior knowledge on user behavior, their attitude change, and the accuracy of bias estimation. Yet, this might be due to the low reliability of self-reported measures. Different levels of actual prior knowledge that we did not capture might have added further noise to our data. The effect of prior knowledge on search behavior, consequences, and metacognitive reflections during search for opinion formation should be investigated in future research.

Lastly, we investigated different factors of user engagement that might be impacted by the interventions, such as their clicking behavior, awareness of bias, and attitude change. However, we did not investigate additional variables that could indicate whether participants thoroughly explored the results (i.e., maximum scroll depth, dwell time), or whether they understood the encountered information (i.e., knowledge gain) and critically evaluated its arguments to form their opinion. Our explorations of data from participants who did not click on any search results revealed, that those participants were just as likely to change their attitude. This observation indicates that the engagement variables captured in these user studies are not sufficient to model search consequences on learning and opinion formation. Future research should investigate searchers' engagement and how it impacts learning and opinion formation more thoroughly, presumably by utilizing both quantitative and qualitative methods.

5.6. Conclusions

We conducted two user studies with the objective of understanding the benefits and risks of behavioral interventions to mitigate users' confirmation bias and support thorough and unbiased information-seeking strategies during search on debated topics. The findings from these studies indicate that obfuscations may risk manipulating behavior rather than guiding it while warning labels without obfuscations effectively encourage users to reduce their interaction with attitude-confirming search results. This suggests that when opting for automatic nudges to decrease cognitive load, users' capacity to actively choose unbiased behavior might be underestimated. We posit that ensuring and facilitating user agency is crucial for interventions that aim at supporting thorough and unbiased information behavior and that in cases where reflective nudging alternatives effectively encourage behavioral change, the risks associated with automatic nudges would not be justified. Obfuscations and potentially other automatic nudging elements to guide search behavior should thus be avoided. Instead, priority should be given to interventions that aim at strengthening human cognitive skills and agency, such as prompting reflective choice to engage with diverse viewpoints. This likely applies beyond our study context, extending to other nudging scenarios that can carry substantial consequences for individuals or society, in which determining what constitutes *beneficial behavior* (i.e., the target behavior towards which users should be nudged) is non-trivial due to complex context and user dependencies.

III

Empowering the Searcher

When reflecting on the dependencies on individuals' epistemic states that we discovered in Part I, it became evident that supporting responsible opinion formation in web search requires not only unbiased search behavior. It also requires diligent behavior, such as actively gathering, evaluating, and comparing information on various arguments to gain a well-rounded understanding of the topic [177, 260, 320]. Attempting to avoid some of the limitations of nudging interventions, namely the risk of harming user autonomy as identified in Part II and failure to address the challenges related to search on debated topics more comprehensively and sustainably, we directed our research focus towards *empowering the searcher* with boosting interventions. *Boosting interventions* aim to cultivate enduring user competencies and should retain their effect even after being presented to the individual [132, 210]. To boost searchers' competencies for unbiased and diligent search on debate topics, we turn towards *intellectual humility* (**IH**)—the competency to recognize the limits of one's knowledge and be aware of the fallibility of one's opinions and beliefs [267]. Part III hence focuses on addressing the following research question:

RQ_{III}: Can we empower individuals to engage in unbiased, as well as diligent search behavior, with interventions that boost their intellectual humility?

In Chapter 6, we reflect on cognitive biases in the broader information search process and confirmation bias specifically and contrast nudging and boosting interventions, based on a review of literature. Inspired by the insights, we suggested a transition from interventions that directly guide search behavior towards interventions that modify behavior indirectly by boosting searchers' intellectual humility. For this transition, we outlined research questions and potential research challenges. In Chapter 7 we present a user study on IH boosting interventions. For that, we identified three interventions that effectively boost self-reported IH. We then tested their effect on the search behavior of strongly opinionated individuals in a familiar search environment. Additionally, we explored the broader role of IH in the search process. We did not find effects on strongly opinionated searchers' behavior. Yet, explorations revealed that searchers' level of intellectual humility is relevant for rewarding search experiences, but that IH boosting interventions require a more direct integration in the search process to disrupt behavioral patterns in familiar information environments. Further, they would likely be more effective in achieving unbiased and diligent search behavior when combined with increased viewpoint-ranking transparency, which we identified as a valuable measure in Part I.

6

Searching for the Whole Truth: Harnessing the Power of Intellectual Humility to Boost Better Search on Debated Topics

This chapter is based on a late-breaking work paper: **Alisa Rieger**, Frank Bredius, Nava Tintarev, and Maria Soledad Pera. "Searching for the Whole Truth: Harnessing the Power of Intellectual Humility to Boost Better Search on Debated Topics". In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems.* CHI EA '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 1–8. ISBN: 978-1-4503-9422-2. DOI: 10.1145/3544549.3585693

6.1. INTRODUCTION AND MOTIVATION

Search engines are widely used to locate information–from a simple definition of a word to resources that are then used to inform searchers' opinions or gather advice before making a decision on debated topics, issues of ongoing discussion, such as *should people become vegan?* (see Figure 6.1). Regardless of the aim of the search, according to *Kuhlthau's* model, the information search process **(ISP)** involves six stages: *Task initiation* (e.g., reason to search for information), *topic selection* (e.g., setting a search goal), *pre-focus exploration* and *focus formation* (e.g., querying and browsing the results), *information collection* (e.g., engaging with search results), and *search closure* (e.g., analysing and synthesizing information) [185]. When focused specifically on debated topics, we posit that the ISP is particularly cognitively demanding since it involves issuing unbiased queries, browsing vast amounts of retrieved resources, learning about an often complex subject matter, accepting a certain level of uncertainty, and thinking critically to objectively assess information, even if the topic is emotionally charged and might pose a threat to personal values. Consequently, a user might experience cognitive biases, such as confirmation bias – the tendency to favor information that confirms prior attitudes, beliefs, and values [244], during the search process [18, 87, 134, 263, 271]. This can impede the user from making well-informed decisions [263, 351], and in a societal context, be a source of increasing extremism and polarization [134, 202].

6

Existing confirmation bias mitigation approaches nudge users towards engagement with attitude-opposing viewpoints, such as preference-inconsistent recommendations [318], alternative query suggestions [270], and warning labels and obfuscations of attitude-confirming search results [298]. These interventions are crucial first steps toward addressing the risks of confirmation bias during web search. However, the interventions themselves cause *undesired side-effects*: Nudging strategies that tap into the *automatic* mind by modifying the ease of accessing some information (friction) harm user autonomy [49, 119, 210], cause reactance such as decreased exploration [49, 270], prevent the detection of incorrect applications [298], and can trigger a feeling of being censored [369]. Thus, such nudging approaches are criticized for the risk of paternalism, enabling manipulation with malicious intentions (e.g., censoring), and the lack of learning [119, 157, 210]. Further, these interventions focus on specific search behaviors (e.g. clicks on attitude-opposing search results). This scope is often too narrow and does not capture the full complexity of the broader problem: Confirmation bias can impact the users' *search behavior* during the whole search process, from querying to assessing and remembering arguments made in the retrieved documents [18, 372]. Consequently, search behavior should be evaluated in a more comprehensive manner, for example by identifying whether users engage in exploratory or lookup search [15], or what search roles (e.g., non-motivated searchers who stop at the first result, confident and competent power searchers) they take on [169, 190] after being exposed to different interventions.

We posit that the value of the aforementioned interventions in real-world settings is somewhat limited when considering the risk of associated side effects. This calls for the need to explore alternative methods that can sustainably guide searchers

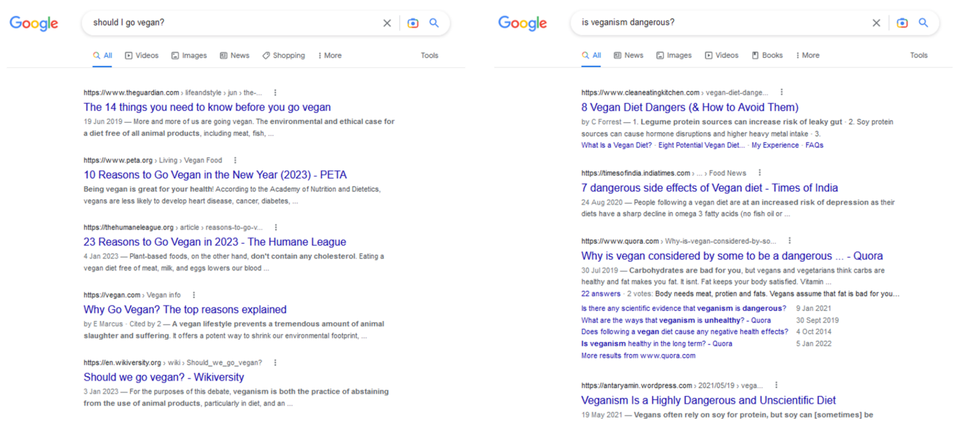


Figure 6.1: **Search engine result pages (SERPs) generated by Google on Jan 18 2023 for queries on debated topics.** Although SERP results are relevant to the respective queries, it is evident that query variations lead to different results: In the SERP for the query *should I go vegan?* (left), most results support veganism; for *is veganism dangerous?* (right) all results oppose veganism.

to responsibly engage with resources on debated topics. We propose to shift confirmation bias mitigation efforts towards enhancing users' *meta-cognitive abilities* that lead to less biased behavior. Particularly vital for confirmation bias mitigation is users' *intellectual humility* **(IH)** [66, 268, 269]. IH describes a metacognitive core consisting of *recognizing the limits of one's knowledge and being aware of the fallibility of one's opinions and beliefs* [267]. Researchers have successfully achieved temporary boosts of IH, for instance, simply by informing participants about the benefits of high IH [267, 269]. This is yet to be explored in the context of web search. Boosting IH to indirectly modify user behavior instead of nudging user behavior directly would avoid harming users' autonomy, be less prone to abuse and errors, and tackle the risks of confirmation bias and other factors that impede good search behavior more comprehensively and sustainably [132, 178, 210].

In this chapter, we approach the following research questions:

RQ1: How can we shift towards interventions that prioritize user autonomy when supporting searchers in navigating web search on debated topics?

With this work, we present and motivate our vision of autonomy-preserving interventions to support better search behavior. As captured in Figure 6.2, interventions boosting users' IH *(meta-cognitive state)* could empower users to explore and engage with different viewpoints *(search behavior)*, critically assess the encountered information when forming opinions and making decisions (*search consequences*), and ultimately contribute to a more inclusive and tolerant society where search engines equip and encourage individuals to gain well-rounded knowledge on debated topics before forming opinions or making decisions.

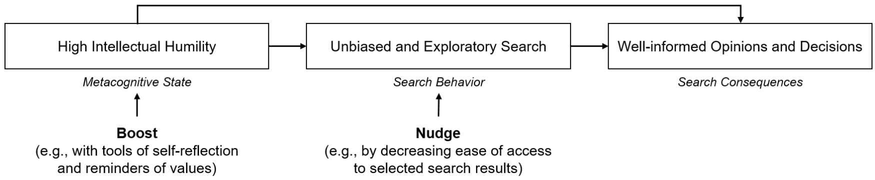


Figure 6.2: **Boosting and nudging for well-informed opinions and decisions after searching on debated topics.** Boosting approaches that target metacognitive states could increase users' intellectual humility by fostering their existing competencies and motivations. Boosting the state of IH could indirectly modify users' search behavior (e.g., increased exploration of diverse viewpoints) and search consequences (e.g., opinions based on more accurate evaluation of strength of arguments), differing from the direct manipulation of users' search behavior of existing nudging approaches.

6.2. RELATED WORK

In this section, we discuss background literature concerning different areas of research that form the basis for the proposed vision.

6.2.1. DEBATED TOPICS AND THE ISP

Debated (controversial/disputed) topics are subjects of an ongoing discussion on which individuals or groups do not have the same opinions. While some debated topics are extremely one-sided and supported by solid scientific evidence, others are less settled because there are reasonable arguments on either side. People use search engines to find information on debated topics, for example, to form opinions or gather advice before making a decision [50, 370]. This would imply that the objectives for this search task are (1) gaining a well-rounded understanding of the topic and the different arguments by gathering information and (2) forming opinions or making decisions in response to all the collected information.

Achieving these objectives requires *exploratory* search that is focused on investigation and learning, as opposed to *lookup* search aimed at retrieving facts to answer specific questions [215]. Exploratory search activities are complex and perceived as challenging since they require users with different needs and abilities to critically analyze, synthesize, and evaluate information [377]. Moreover, searchers can experience a certain level of threat when seeking resources on debated topics: Strong prior beliefs and convictions, as well as political, religious, or ethical values, are likely to be challenged by attitude-opposing viewpoints [269].

According to Kuhlthau's ISP model [185], searching for information involves six stages. The particularly high level of cognitive demand imposed on searchers becomes apparent when contemplating what these stages encompass for web search on debated topics.

1 *Task initiation* is often prompted when users need to form their attitude on a topic or seek advice before making a decision. Such situations can increase the stress perceived during the ISP, e.g., through time pressure, uncertainty, and the complexity of the subject matter of many debated topics [263].

2 *Topic selection* compels users to approach the search task with the goal of

gaining knowledge and being well-informed about the subject matter instead of finding evidence to support their prior attitude.

3 & 4 *Pre-focus exploration and focus formation* require users to formulate unbiased queries, and navigate and accurately assess the quality of vast amounts of resources.

5 *Information collection* takes place as the users engage with and evaluate different viewpoints and arguments to extend knowledge on and gain a well-rounded understanding of the often complex subject matter.

6 *Search closure* results from users thinking critically to objectively assess the information they have encountered to be able to make well-informed decisions, even if the topic is emotionally charged and alternative viewpoints threaten users' prior beliefs, convictions, and values.

To deal with the complexity, cognitive demand, and uncertainty of web search for resources on debated topics, users are likely to apply strategies to simplify the search task [18]. A dominant strategy to limit the uncertainty and amount of possible resources to engage with is the human tendency to prioritize information that confirms prior attitudes when searching for, engaging with, and assessing information *(confirmation bias)* [244].

Confirmation bias during search on debated topics can occur at different stages of the ISP, e.g., when employing positive test strategies when querying (focus formation, see Figure 6.1), clicking primarily on attitude-confirming search results (information collection), and actively disregarding information that opposes users' prior attitudes when assessing arguments to form an opinion or make a decision (search closure) [18, 360, 372]. Thus, confirmation bias, in this case, would either limit exploratory search behavior to merely one-sided information or prevent all exploration by causing users to engage in lookup search behavior that aims at retrieving facts to support their prior attitude. In addition to confirmation bias, users' exploration behavior of resources on debated topics can be further inhibited by manifold other cognitive biases that can occur during search [18], or by external obstacles, for example, SERPs with viewpoint-biased rankings [74, 82], or interfaces designed to steer behavior for commercial gain [178]. Lacking exploration of different viewpoints has negative consequences on the quality of attitude-forming and decision-making since it prevents users from building well-rounded knowledge [18, 263]. This has been linked to increased polarization and ideological extremism, thus additionally harming the quality of public discourse [134, 202].

In summary, seeking resources about debated topics in the pursuit of well-informedness is a non-trivial undertaking; it requires complex search behavior known to be cognitively taxing throughout the ISP. This high demand makes users vulnerable to cognitive biases, has a detrimental impact on all ISP stages, and hence prevents users from becoming well-informed, which has been linked to increased polarization and extremism and thus should be mitigated.

6.2.2. Intellectual Humility

According to the definition that Leary et al. [195] developed following discussions with an interdisciplinary group, the metacognitive core of *intellectual humility* describes people's *"recognition that a particular personal belief may be fallible, accompanied by an appropriate attentiveness to limitations in the evidentiary basis of that belief and to one's own limitations in containing and evaluating relevant information"*. Porter et al. [267] have synthesized the common thread within different definitions of IH from various fields and suggest that it encompasses (1) recognizing the fallibility of one's beliefs, and (2) recognizing the limits of one's knowledge.

In the context of knowledge and information behavior, IH can counter typical behavioral patterns that searchers tend to exhibit when subjected to confirmation bias. Intellectually humble people are more likely to indulge in increased information seeking and have a high motivation to gain new knowledge [104, 182, 268]. They also spend more time learning about opposing arguments and reading information countering their prior beliefs [40, 67, 268]. High IH enables people to distinguish the strength of different arguments, even those opposing their prior belief [195]. It can also help users overcome external obstacles that impede optimal search behavior, e.g., by decreasing their susceptibility to false and misleading information [40, 175]. However, searchers are less likely to exhibit IH when engaging with a topic on which they have a strong prior opinion, and/or when their political, religious, or ethical values appear to be challenged [183].

6

IH has been identified as a relatively stable trait (a person's general level of IH). Yet, researchers have observed substantial and systematic within-person variability of IH as a state (a person's level of IH in a specific context) [267, 391]. IH on the trait level positively correlates with other user traits, such as the need for cognition and cognitive reflection [182, 195, 268], or open-minded thinking and curiosity [183, 195, 399]. Further, people exhibiting behavior related to high IH are likely influenced by their cultural background. Someone living in an environment that requires high social coordination is more likely to be intellectually humble than someone who lives in an individualistic environment [109].

Researchers have developed several methods to measure IH which differ in type (questionnaire vs. behavioral task), the aspects of IH they emphasize (limits of knowledge, fallibility awareness), whether they measure IH on the trait- or state-level, and the assumed dimensionality of IH (up to four) [267]. Alfano et al. [6] developed one of the most extensively tested measures of IH. Their scale captures the trait level of IH on the four dimensions of *open-mindedness* vs. arrogance, *intellectual modesty* vs. vanity, *corrigibility* vs. fragility, and *engagement* vs. boredom. However, questionnaire-based measures of IH on the trait level have been criticized for being vulnerable to *social desirability bias* and for failing to detect context- and intervention-dependent variability [267]. Behavioral task-based measures cannot be distorted by self-report biases. Still, they might only capture a segment of artificial behavior, induced by the experimental setting. Porter et al. [267] suggest applying questionnaire-based measures and asking people to recall a specific situation when filling out the questionnaire or to measure the trait level by repeated measures of the state level of IH to mitigate response bias. To sidestep issues related to questionnaire- and behavioral task-based measures, Christen et al. [58] have investigated an indirect method of assessing IH with natural language processing **(NLP)** techniques to extract different dimensions of IH from written text.

Summarizing the findings on IH, this metacognitive concept encompasses recognizing the fallibility of one's beliefs and the limits of one's knowledge. High IH was found to counter behavioral patterns during information seeking that are common for confirmation bias. Substantial within-person variability of IH suggests that it can also be considered a context-dependent state. Researchers have developed many methods to measure IH on the trait- and state level, such as multidimensional questionnaires or behavioral tasks, which have different advantages and disadvantages.

6.2.3. Mitigating Cognitive Bias during Search

To support better search behavior, enabling well-informed attitude-forming and decision-making, Lorenz-Spreen et al. [210] propose *effective web governance* through the application of behavioral interventions in form of nudging or boosting. This is in line with the guide to cognitive debiasing by Soll et al. [326], who suggest either to *modify the environment* or to *modify the user*. So far, approaches to mitigate confirmation bias during web search have focused on *modifying the environment*, e.g., with preference-inconsistent recommendations [318], alternative query suggestions [270], and warning labels and obfuscations of attitude-confirming search results [298] to nudge searchers towards increased engagement with attitude-opposing viewpoints.

Interventions that nudge user behavior by modifying the search environment and ease of access to different search results are non-transparent and target automatic thinking processes that can *harm user autonomy* [49, 119]. This indicates that the decision of what information users engage with is, without users' awareness, not entirely theirs. This non-transparency can result in users being unable to detect incorrect applications of the nudge [298], enables concealed applications of the approach with malicious intentions, or triggers a feeling of being censored [369].

Boosting interventions that *modify the user*, generally preserve user autonomy by aiming at fostering people's existing cognitive or motivational competencies, thus encompassing a learning component and, unlike nudging, remaining effective even after the intervention [132, 210]. A promising metacognitive concept for mitigating confirmation bias is IH, the variability of which, as a context-dependent state, generates opportunities for interventions that attempt to boost it. Researchers have explored different strategies that temporarily boost IH on the state-level [267]. For instance, asking participants to reflect on scenarios from a self-distanced perspective [109], quizzing participants on a topic to make them realize the limits of their knowledge [153, 222], or simply informing them about the benefits of IH [269]. Krumrei-Mancuso and Newman [183] observed that approaches to boost IH, such as asking participants to complete a short IH scale, might require personalization to be effective for all users. The authors found that priming IH increased responsiveness to information on a debated topic in high IH users while for low IH users, the priming did not have an effect.

In summary, while effective, existing approaches against confirmation bias during search that apply nudges to directly modify user behavior, run the risk of harming user autonomy. This motivates our quest for alternative interventions that aim at boosting users' metacognitive states. Boosting approaches that aim at fostering metacognitive concepts could be an autonomy-preserving alternative. For that, IH is a particularly relevant concept since high IH on the trait level was shown to counter confirmation bias during different stages of the information-seeking and attitude-forming process. The substantial within-person variability of IH as a context-dependent state generates opportunities to boost IH, which have been applied successfully in non-search contexts.

6.3. Vision: Boosting IH to Mitigate Confirmation Bias

In the context of bias mitigation, non-transparent and automatic nudging approaches can be seen as a form of paternalism, as they suggest that users' behavior is faulty and requires (manipulative) correction. Boosting approaches that target metacognitive skills, on the other hand, view users as individuals who carry the competencies and motivation needed to overcome their biases within them and, from that, can develop them [210]. To avoid the risk of harming user autonomy, we propose developing interventions that boost the metacognitive state of searchers (see Figure 6.2). Such boosts could positively impact users' overall search behavior and the search consequences and mitigate their confirmation bias throughout the ISP, from setting the search goal (topic selection) to synthesizing the encountered information (search closure).

Research Questions. Motivated by gaps observed in existing literature (Section 6.2) and focused on setting a research foundation on the impact of IH on search on debated topics with a specific focus on confirmation bias, we outline an initial set of research questions: 1-4 focus on effective boosting interventions and their impact on search behavior and opinion formation; the remaining RQs guide a wider research scope encompassing personalization, long-term effects, other search tasks, and alternative search paradigms.

(1) What are effective interventions to boost IH in web searchers?

(2) Does boosting IH modify users' search behavior when searching for resources on debated topics?

(3) Does boosting IH mitigate confirmation bias during search on debated topics?

(4) Does boosting IH enable better-informed opinion-forming and decision-making in users?

(5) How would the boosting interventions need to be personalized to be effective for users with different characteristics and abilities?

(6) What are the longer-term effects of boosting IH on search behavior and search consequences?

(7) How would the boosting interventions need to be adapted to support better search behavior and consequences for search tasks beyond debated topics?

(8) How would the boosting interventions need to be adapted to be effective in non-textual search paradigms (e.g. with conversational agents or chatbots)?

Challenges. Certain challenges will be encountered when planning how to evaluate the applicability and effectiveness of boosting approaches against confirmation bias during web search for resources on debated topics.

Categorizing Search Behavior. We are interested in the effects of boosting interventions on the quality of users' overall search behavior throughout the ISP. For this, we need to operationalize and map behavioral variables (queries, clicks, time spent, etc.) to categorize search behavior (e.g., exploratory vs. lookup, viewpoint biased vs. unbiased, focused vs. unfocused, or motivated vs. unmotivated). This mapping can be informed by prior attempts, for example, identifying behavioral indicators of exploratory vs. lookup search behavior [15], or mapping quantitative indicators of search behavior to different search roles that users take on [169, 190]. It would also be interesting to look at the general relationship between IH and search roles by investigating whether IH affects people differently who tend to take on different search roles.

Measuring IH. To test the degree to which interventions lead to increased IH, we need to devise a method to assess the change in the participants' IH from prior- to post-boosting. However, applying a questionnaire might boost IH itself and thus distort the results. Alternatively, the effect of interventions on IH could be investigated in laboratory studies to capture behavioral metrics, e.g., with NLP- or eye-tracking-based measures of IH [29, 58].

Evaluating Search Consequences. Ultimately, we aim for simple and applicable interventions that positively impact the consequences of web searches on debated topics, namely by achieving better-informed attitude-forming and decision-making. While informedness can be measured with knowledge questionnaires on the topic, the responsiveness of a decision or attitude to the information someone has is very challenging. This could be evaluated with an artificial topic in a very controlled setting with a designated *"correct"* attitude participants should have after the search sessions. However, such a setting would likely fail to reflect the complexity of search on debated topics and thus lack ecological validity.

Creating Realistic Search Scenarios. To ensure the validity of potential findings, the search scenario and environment need to assimilate real-world search on debated topics. This requires creating a sound task initiation, recreating a search environment that initiates confirmation bias and allows for a large range of possible search interactions while having access to information about the search results' viewpoints to evaluate bias.

6.4. CONCLUSIONS

With the work we presented in this chapter, we call for a shift in how we attempt to tackle the manifold and complex challenges inherent to search on debated topics. In their majority, approaches so far have focused on directly nudging users' behavior. Instead, we propose to modify user behavior indirectly by boosting users' metacognitive state of intellectual humility. Doing so would avoid the risk of harming user autonomy and address the broader issues related to search on controversial topics in a more comprehensive and sustainable manner. To make a meaningful impact, it is crucial to take a holistic approach to problem-solving, driven by interdisciplinary efforts to identify efficient boosting approaches, develop the technical aspects of integrating boosting interventions into search engines, design an easy-to-use interface that responds to different users, and evaluate the individual and societal impact.

7

From Potential to Practice: Intellectual Humility during Search on Debated Topics

This chapter is based on a full conference paper: **Alisa Rieger**, Frank Bredius, Mariët Theune, and Maria Soledad Pera. "From Potential to Practice: Intellectual Humility During Search on Debated Topics". In: *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*. CHIIR '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 130–141. DOI: 10.1145/3627508.3638306

7.1. Introduction

The considerable cognitive effort and diligence required to engage with information on debated topics, coupled with an information environment of opaque algorithmic curation that evokes over-reliance as observed in web search, poses an obstacle to informed opinion formation or decision-making [226]. This is compounded by the fact that, in general, searchers' information-seeking habits are known to be shaped by various individual characteristics [46, 229, 267, 362], emphasizing that there is likely no one-size-fits-all solution for fostering search behavior that leads to informedness on debated topics. This leads us to question *how we can empower individuals to overcome the challenges associated with web search on debated topics, ultimately engaging in unbiased, as well as diligent search behavior.*

As a starting point for answering this question, we turn to **intellectual humility (IH)**, a key characteristic for unbiased and diligent information seeking that pertains to the awareness of one's own epistemic limitations, i.e., the limits of one's knowledge and fallibility of one's beliefs [66, 267]. Individuals with high IH generally have a high motivation to seek information and gain knowledge [182, 268]. They tend to spend more time learning about attitude-opposing arguments and can better identify the strength of different arguments [40], making them less prone to biased behavior when engaging with information on debated topics. A number of promising approaches to boost IH, such as brief reflection exercises [181], reading about the benefits of IH [269], or reading about the plasticity of intelligence [268], have emerged. In light of these discoveries, researchers see great potential in IH boosts to improve the quality of opinions and decisions at the individual level, as well as foster more harmonious intergroup relationships and reduce polarization at the societal level [267, 295]. Up to this point, however, approaches to boost IH have primarily been assessed in terms of their impact on self-reported IH and reflection tasks, rather than their influence on actual behavior within a familiar information environment.

To determine whether the potential of IH would translate into unbiased, diligent search habits on debated topics in practice, we conducted a preregistered user study with 299 participants. To control scope, we center this study on opinionated searchers (i.e., reporting moderate and strong attitudes) who were found to be least open to processing attitude-opposing information [358] and thus in greater need of support for unbiased search. Guided by three preregistered (RQ1, RQ2, RQ3) and one exploratory research questions (RQe), this study investigated (1) the effects of three interventions of varying complexity that we found to boost self-reported IH in a pre-study, detailed in Section 7.1 (*prime, remind, reinforce*, see Figure 7.1), and (2) the role of IH during search on debated topics more broadly.

RQ1 Do the interventions of *prime*, *remind*, and *reinforce* that boost intellectual humility lead to **decreased confirmation bias** during search result selection on debated topics?

RQ2 Do the interventions of *prime*, *remind*, and *reinforce* that boost intellectual humility lead to **increased search diligence** during search on debated topics?

7

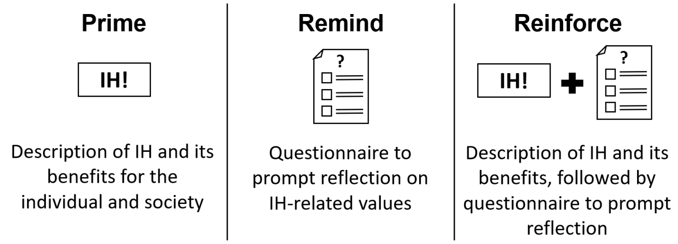


Figure 7.1: **Boosting interventions inspired by the concept of Intellectual Humility (IH).**

RQ3 Are there **differences between the effects of the interventions** *prime*, *remind*, and *reinforce* on search diligence and confirmation bias during search on debated topics?

RQe How does **IH factor into the broader search process** of opinionated individuals who conduct searches on debated topics?

We probed the effects of the interventions with a between-subjects design. We exposed participants to one of the three boosting or two control conditions (see Figure 7.3). They then used a mock search engine result page **(SERP)**—resembling a familiar web search interface—to learn more about a debated topic on which they reported to have a strong attitude. During the search task, we logged their interactions with the SERP. To explore how IH factors into the broader search process, we asked participants to report their attitude, perceived knowledge gain, rationales for their behavior, and reflections on the search task after they finished the search.

Analysis of participants' search behavior did not corroborate that either of the three interventions affected participants' confirmation bias and search diligence. In other words, the interventions that we empirically found to boost self-reported IH in the pre-study, could not empower opinionated individuals to overcome the challenges of web search on debated topics in practice. We attribute the lack of differences in search behavior between control and boosting conditions to (i) investigating effects on actual search behavior in a familiar environment that might diminish the effect of the interventions by leading users to resort to default behavior such as relying on the ranking; and (ii) targeting opinionated searchers who might be less inclined to display IH in their actions, even though the boost affected their self-reported IH.

Based on the insights emerging from exploring the role of IH during the broader search process, we deduce that even though the boosting interventions could not successfully change behavior, IH should still be considered as a lever in the pursuit of promoting unbiased and diligent search on debated topics. While we did not observe direct links between the level of IH and search behavior, searchers' reflections on the search task suggest that those with high compared to low IH might approach searching on debated topics with greater ease and perceive to gain more knowledge. Further, our explorations indicated that searchers who reported having approached the search in an IH-driven way were more likely to exhibit search diligence than those who relied on the ranking. These findings lead us to argue that for IH-boosting approaches to cause behavioral change in familiar search environments, they likely need to target searchers' motivation to approach the search task in an IH-driven way and to be combined with interventions that are more directly integrated into the search process and induce appropriate reliance on the search system and ranking, e.g., epistemic cues [210].

With this chapter, we advance the understanding of the potential of IH-boosts which hold the promise of being a remedy for various epistemic societal challenges. It does so by taking an initial step towards testing such interventions in practice with a preregistered user study (N=299), empirically testing the effect of interventions that boost self-reported IH on behavior during web search on debated topics in a familiar search environment. It further contributes to the understanding of the role of IH in the broader process of search on debated topics, encompassing attitude change, perceived knowledge gain, and searchers' reflections on their search behavior and the search task, alongside search behavior. We did not find evidence for and impact of the IH-boosts on search behavior. Regarding the role of IH in the search process, we found that while IH shapes task load, IH-related motivations in approaching the search task rather than IH directly shape search behavior in this context. Based on our insights, we present **design implications** for interventions that aim at supporting unbiased and diligent search behavior and **methodological implications** for research efforts that aim at empowering individuals online. In the pursuit of open science, we made the preregistration with detailed descriptions of the study plan and dataset with questionnaire responses and behavioral data from search logs publicly available.[1]

7.2. Related Work

Most search engines lack support for the activities searchers need to carry out to satisfy complex information needs [206, 215, 320, 325]. For instance, mainstream search engines may fail to support diverse search intentions, information-seeking strategies, and transitions between them [206, 320]. Further, the opaque algorithmic curation of results that are displayed to the searcher makes it difficult to understand the information space and recognize whether sufficient information has been gathered to conclude the search [178, 226]. However, there have been recent calls to improve support for such complex search tasks, e.g., providing transparency over the ranking, displaying meta-information alongside the search results, or visualizing search intents and the information space [307, 320, 325].

To *empower individuals* to navigate online environments, Lorenz-Spreen et al. [210] propose behavioral interventions such as *nudging* and *boosting*. Nudging aims at steering user behavior by altering the choice architecture (e.g., altering the effort required to access or evaluate selected information, setting defaults) [49, 337]. In contrast, boosting interventions aim at fostering user competencies that facilitate navigating online environments and, unlike nudging, offer the advantage of

[1] https://osf.io/ktysd/?view_only=e9d8e67f568f41559edf277b4c2645cc

upholding user autonomy, as well as remaining effective over an extended period of time [132, 210]. For search on debated topics, researchers have used nudging approaches to encourage individuals to explore diverse viewpoints to facilitate informed opinion formation and decision-making. For instance, obfuscations with warning labels of attitude-confirming search results [298], displaying labels that indicate the stance of the search results [382], or tag clouds that reveal experts evaluations [317] were found to reduce confirmation bias, while query priming was found to promote diligent search behavior and increased exploration [384]. Researchers have also investigated *argument retrieval* systems that can facilitate web search on debated topics by directly presenting distinct arguments retrieved from search results [3, 38, 366]. Boosting interventions, promising to empower users to navigate other online challenges, such as microtargeting [209], have not yet been investigated in the context of search on debated topics.

Since general information-seeking behavior is known to be shaped by various user characteristics [46, 229, 267, 362], the challenges associated with search on debated likely *do not affect all individuals equally*. Some characteristics that were found to affect search are context-dependent (e.g., attitude strength [358]), and others are more stable (e.g., the need for cognition—an individual's general tendency to organize their experience meaningfully [46, 362]). For example, searchers who have a strong compared to weak prior attitude on the topic they search on were observed to be less open to processing attitude-opposing information [358], and individuals with a low compared to high *need for cognition* were observed to be less diligent searchers [362]. This highlights that heterogeneous searchers have varying requirements when it comes to supporting unbiased and diligent search on debated topics.

The central element investigated with this study, *IH*, is linked to different cognitive, social, and personality traits that shape information behavior [66, 267]. It entails recognizing the limits of one's knowledge and being aware of the fallibility of one's beliefs. IH can be measured as a context-dependent user state (i.e., an individual's degree of IH in a specific context) and a stable user trait (i.e., an individual's general degree of IH) [6, 141, 267]. High IH was found to reduce the propensity for patterns that indicate biased and non-diligent information-seeking behavior, e.g., limited curiosity, low intrinsic motivation, and eagerness to invest effort in learning [182, 269], as well as little engagement with opposing viewpoints [40]. Looking at societal challenges arising from these information-seeking patterns, high IH was linked to reduced hostility towards individuals with opposing views [328], decreased affective polarization [40], and diminished susceptibility to misinformation [175]. In light of these observations, researchers see potential in interventions that boost IH to function as an antidote to such epistemic societal challenges [104, 267]. While simple approaches that effectively boost self-reported IH have indeed been identified [181, 268, 269], their effect on real-world information behavior, and web search on debated topics, in particular, has yet to be explored.

7.3. Pre-Study: Boosting Interventions

In our quest for a simple and effective intervention that could be practically implemented in a real-world search setting, we considered different boosting approaches that, as identified in the recent research, could foster web users' cognitive competencies.[2] Given the context of our work, we were particularly interested in interventions to boost searchers' intellectual humility by means of *self-reflection* and priming *societal values*. Further, we adopted the approach proposed by Lorenz-Spreen et al. [210] and considered alternatives of varying complexity.

This resulted in three boosting interventions (see Figure 7.1).

1. **Prime:** informing searchers of the societal values related to IH by briefly describing the concept and its benefits;

2. **Remind:** raising searchers' awareness of IH and reminding them of their own values related to it by asking them to fill in the *multidimensional IH scale* [6];

3. **Reinforce:** reinforcing values by reminding searchers of societal values before reminding them of their own values by briefly describing IH and its benefits and subsequently asking them to fill in the *multidimensional IH scale* [6].

To test whether interventions *effectively* boost IH, we conducted a between-subjects pre-study approved by the ethics committee of our institution. We recruited 251 participants via *Prolific* [277], of whom 240 passed the attention checks and were included in the analysis.

Procedure. We asked participants to report their attitude on all nine debated topics featured in the dataset with viewpoint-annotated search results by Draws et al. [71] by reporting their agreement with a statement on each topic (e.g., *Is drinking milk healthy for humans?*) on a seven-point Likert scale ranging from *strongly disagree* to *strongly agree*. We used these topics as we sourced the search results presented to the participants in the main user study from this dataset. Participants were assigned to a debated topic on which they reported having a strong attitude (i.e., *strongly disagree/agree*, or *disagree/agree*). We then measured participants' context-dependent, self-reported IH with the *Specific Intellectual Humility Scale* (seven-point Likert scale) [141]. We formulated the questionnaire items in the context of the assigned topic (e.g., *My views about TOPIC are just as likely to be wrong as other views.*). Subsequently, we randomly assigned and exposed them to one of the three boosting interventions (prime, remind, reinforce) or the control intervention (ATI control). In the ATI control intervention, participants were asked to fill out the *Affinity for Technology Interaction* (ATI) scale. To conclude the task, we asked the participants to answer the questions of the *Specific Intellectual Humility Scale* once more and calculated the difference to their initial IH score.

Results. An ANOVA revealed evidence for a moderate effect of the intervention type on self-reported IH difference ($F(3,236) = 5.99, p < .001, f = 0.28$). As expected,

[2]For an overview of different digital boosting approaches, see https://www.scienceofboosting.org/tag/digital/

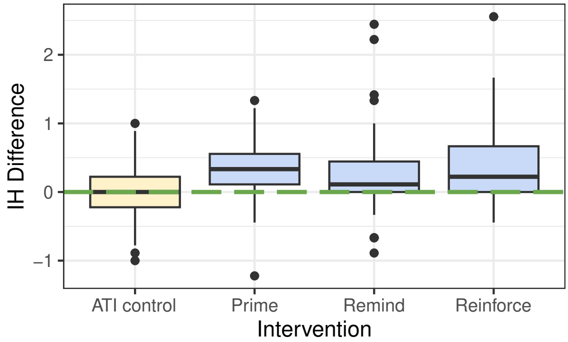


Figure 7.2: **Prestudy: IH difference per intervention condition.** Boxplots illustrating the distribution of the difference in IH levels across participants per intervention condition. The boxplots show medians and quartiles of differences in IH levels, measured with the IH questionnaire before and after participants' exposure to the selected intervention.

IH of participants in the ATI control condition did not change ($mean = 0.01, SE = 0.05$), while IH of participants in the prime ($mean = 0.33, SE = 0.06$), remind ($mean = 0.2, SE = 0.08$), and reinforce ($mean = 0.37, SE = 0.07$) conditions increased (see Figure 7.2). Based on these findings, we tested the effect of all three boosting interventions on search behavior in the main user study.

7.4. USER STUDY METHODOLOGY

To investigate the three preregistered RQs (see Section 7.1), we tested the following hypotheses with a randomized controlled trial between-subjects design:

- **H1 (confirmation bias):** Searchers who are exposed to an intervention that boosts IH click less on attitude-confirming search results than other searchers.
- **H2 (search diligence[3]):** Searchers who are exposed to an intervention that boosts IH
  - **a:** click on lower-ranked documents than other searchers.
  - **b:** display longer dwell time than other searchers.
  - **c:** spend more time on the search task than other searchers.
  - **d:** make more clicks than other searchers.
- **H3 (differences):** The *reinforce* boosting intervention will have a stronger effect on users' search behavior than the *remind* and *prime* boosting interventions.

To address the exploratory RQ, we investigated the effects of *measured IH* on search behavior, attitude change, and self-reported knowledge gain. Further, we

[3]During preregistration, we employed the term *search effort* rather than *diligence*. However, due to the potential ambiguity associated with *effort* in the context of web search, we opted for *diligence* as it more accurately conveys our intended meaning.

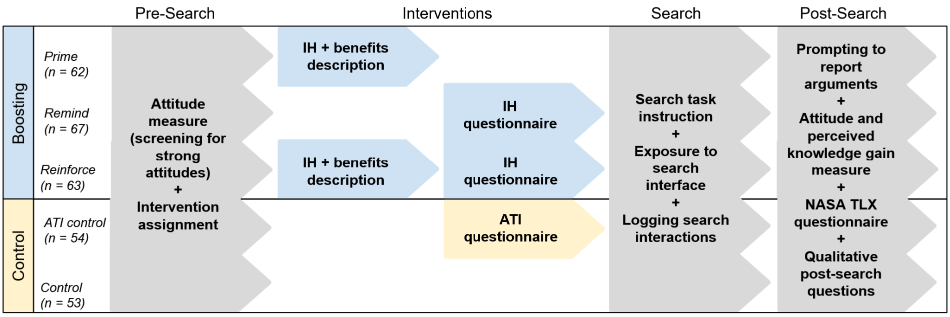


Figure 7.3: **Procedure of the user study per intervention condition.**

investigate the rationales that participants reported for their behavior, whether they align with observed search behavior, and whether they are related to attitude change and self-reported knowledge gain. Finally, we explore links between participants' reflections on the search task, their search behavior, level of IH, attitude change, self-reported knowledge gain, as well as their reported rationales for their behavior.

Comprehensive descriptions and motivations regarding the study's methodology, materials (such as search results and questionnaires), as well as hypotheses and analysis plan, can be found with the preregistration linked in Footnote 1.[4]

7.4.1. Procedure

We recruited participants via *Prolific* and used *Qualtrics* [283] for pre- and post-search questionnaires. We collected the data for this user study with the following procedure (see Figure 7.3), approved by the ethics committee of our institution.

- **Participant screening.** In a designated task, we asked crowdworkers to report their agreement with a statement on each of the aforementioned nine topics, using a seven-point Likert scale ranging from *strongly disagree* to *strongly agree*. Subsequently, we excluded topics for which only a few participants reported having a strong attitude (i.e., *strongly disagree/agree* or *disagree/agree*), resulting in the following set of six topics: *Should people become vegetarian? Is drinking milk healthy for humans? Should students have to wear school uniforms? Is homework beneficial? Is obesity a disease? Should bottled water be banned?* Individuals who reported having strong attitudes on three or more of the topics were invited to participate in the study.

- **Pre-search.** After consenting to participate in the study, participants were randomly assigned to one of five intervention conditions (control, ATI control, prime, remind, reinforce, see Figure 7.3). We tested two control conditions, one without any intervention and one in which participants were asked to fill out

[4]We initially planned to investigate the interventions' effects on later search sessions (RQ4 in the preregistration); given the lack of differences in the initial session, we did not proceed with further data collection.

the ATI questionnaire. By including the control condition without intervention we could compare search behavior to that in a standard search setting. In addition, we included the ATI control condition to be able to distinguish whether potential effects of the IH boosting interventions on search behavior can be attributed to boosted IH, or if they might simply be a result of the reflective moment, filling a questionnaire before starting the search task.

Participants were exposed to the assigned intervention before advancing to the following instructions for an open-ended search task:

`A friend is telling you about a discussion they had with a colleague about TOPIC. The conversation made you curious. To learn more about TOPIC, you have decided to conduct a web search.`

- **Search.** We presented the search task instructions for a topic on which the participant reported having a strong opinion during the screening task. From the task instructions, they could advance to the mock search interface, designed to mimic familiar web search interfaces. On the search interface, they could enter a query. If the query passed a similarity criterion to the topic statement, searchers were presented with viewpoint-annotated search results sourced from the dataset by Draws et al. [71]. Per SERP, we displayed ten search results with alternating viewpoints of *supporting*, *opposing*, or *neutral* with respect to the searcher's attitude. We distributed participants equally among conditions with either an attitude-confirming, an attitude-opposing, or a neutral search result on the top rank to control for potential ranking effects on participants' search behavior. Participants could click on the links to retrieve the documents as they would on a common SERP. We logged search interactions with the *LogUI* framework [217].

- **Post-search.** Once participants finished searching, we asked them to report the arguments they encountered to convey a sense of having completed the task. Participants were asked to state their attitude on the topic once more to compute their attitude change, following the method applied in prior research on attitude change in web search (e.g., [74, 82, 297]). In addition, we asked them to report the level of perceived knowledge gain over the search session on a five-point Likert scale ranging from *no knowledge gain* to *substantial knowledge gain*. We then asked participants to fill the NASA task load (NASA-TLX) questionnaire [123], omitting the question on physical demand since the task did not involve physical extortion. Lastly, we invited them to reflect on their behavior (*What made you decide to click on the search results you clicked on?*) and give us feedback on the task.

7.4.2. VARIABLES

In Table 7.1, we describe the variables used in our study to capture the effect of the intervention on searchers' level of confirmation bias during search result selection (**RQ1, H1**, attitude-confirming clicks) and search diligence (**RQ2, H2a - H2d**, lowest

Table 7.1: **Study Variables.** IV, DV, and EV for independent, dependent, and exploratory variables, respectively.

Type	Name	Description
IV	Intervention	The intervention to which participants were exposed prior to the search. One of control, ATI control, prime, remind, reinforce.
DV	Attitude confirming clicks	Proportion of clicks on attitude confirming search results. (H1, H3)
	Lowest rank clicked	Lowest rank of a link that the participant clicked on. (H2a, H3)
	Dwell time	The average time a participant spends on a clicked document in seconds. (H2b, H3)
	Task completion time	The time a participant spends on the search task in seconds. (H2c, H3)
	Cumulative clicks	A participant's number of clicks on unique search results. (H2d, H3)
EV	Intellectual Humility	Score of IH according to responses to the IH questionnaire (only captured for n = 130 in remind and reinforce conditions. Values ranging from 1 to 7.)
	Ranking	Stance of the search result displayed on the top rank. One of attitude-confirming, attitude-opposing, or neutral.
	Topic	Topic assigned to participant. One out of drinking milk, homework, obesity, bottled water, vegetarianism, school uniforms.
	Rationale for behavior	Reported rationale for participants' search behavior (free text categorized into one of driven by IH, ranking, confirmation bias, content/form, task/unclear).
	Attitude change	Difference between pre- and post-search attitude. Positive values indicate a strengthening, and negative values a weakening of the initial attitude.
	Knowledge gain	Self-reported knowledge gain for the topic searched on.
	Reflection on search task	NASA-TLX results, perceived levels of mental demand, temporal demand, performance, effort, and frustration. Values range from 0 to 100.

rank clicked, dwell time, task completion time, cumulative clicks)[5], as well as to determine differences in search behavior across the five interventions (**RQ3, H3**). To investigate how IH factors into the broader search process (**RQe**), we considered exploratory variables beyond search behavior. Details on how we captured the different variables are outlined in Section 7.4.1 and the preregistration linked in Footnote 1. Lastly, we collected data on participants' age and gender to provide contextual information about the study sample.

7.4.3. DESCRIPTION OF THE SAMPLE

With an a priori power analysis (with $f = 0.25$, $\alpha = \frac{0.05}{6} = 0.0083$ (due to testing six hypotheses), $(1-\beta) = 0.8$, and 5 groups (i.e., 5 intervention conditions), we determined a sample size of 285 participants. Initially, 349 participants completed the study, of which 299 met the preregistered inclusion criteria for data analysis (passed attention checks, clicked on at least one search result). Of the 299 participants, 44% reported to be female, 55% male, and the rest non-binary/other. Regarding their age, 37% reported to be between 18 and 25, 35% between 26 and 35, 17% between 36 and 45, 7% between 46 and 55, 3% between 56 and 65, and 1% more than 65 years old. Participation was rewarded with £2.30 (mean = £9.32/h).

7.5. RESULTS

Here, we first present the results of testing the hypotheses on the effect of the boosting interventions on confirmation bias and search diligence (Section 7.5.1).

[5] Collectively, these variables reflect behaviors that demonstrate searchers' commitment to thoroughly exploring, engaging with, and considering various resources and thus approximate search diligence.

Table 7.2: **Confirmation Bias and Search Diligence per Intervention Condition.** Means and standard errors for attitude-confirming clicks, lowest rank clicked, dwell time, task completion time, and cumulative clicks for each intervention condition.

	Attitude confirming clicks		**Lowest rank clicked**		**Dwell time**		**Task completion time**		**Cumulative clicks**	
	mean	*SE*	*mean*	*SE*	*mean*	*SE*	*mean*	*SE*	*mean*	*SE*
Control (n = 53)	0.33	0.05	7.3	0.8	50.1	11.2	303	37	3.1	0.3
ATI control (n = 54)	0.37	0.04	10.9	2.7	45.4	5.8	349	51	4.1	0.4
Prime (n = 62)	0.38	0.04	9.1	1.5	49.7	10.4	319	32	3.5	0.3
Remind (n = 67)	0.32	0.04	8.8	1.7	41.7	5	249	23	3.2	0.2
Reinforce (n = 63)	0.38	0.04	7.7	1.5	42.6	5.5	280	26	3.3	0.3

We then provide an overview of our exploratory findings, aimed at enhancing our understanding of the results from hypothesis testing and addressing the exploratory research question, regarding the role of IH in the broader search process (Section 7.5.2).

7.5.1. HYPOTHESES TESTING

Effect on confirmation bias. Results of an ANOVA indicated no evidence for H1, an effect of the interventions on the proportion of attitude-confirming clicks ($F(4,294) = 0.39, p = .81, f = 0.07$). We explored whether potential topic and ranking effects might have prevented us from seeing differences between control and intervention conditions. While we did not find evidence for topical differences ($F(5,293) = 1.18, p = .32$), we noted that the viewpoint of the top-ranked search result affected the proportion of clicks on attitude-confirming search results ($F(2,296) = 6.68, p = .001, f = 0.21$). Participants who saw a neutral search result on the top rank clicked on a lower proportion of attitude-confirming search results ($mean = 0.26, se = 0.03$) than those who saw an attitude-confirming ($mean = 0.41, se = 0.03$) or attitude-opposing ($mean = 0.38, se = 0.03$) search result (see Figure 7.4). Yet, when controlling for the effect of ranking we still did not find evidence for an effect of the interventions ($F(4,292) = 0.52, p = .72$). Noteworthy, we observed that across the five intervention conditions, the mean proportion of attitude-confirming clicks was between 32.4% and 37.6%, indicating overall low confirmation bias (see Table 7.2). Addressing RQ1, these findings do not substantiate that the boosting interventions decrease searchers' confirmation bias.

Effect on search diligence. The MANOVA results indicated no differences between the intervention conditions for any of the variables indicating search diligence ($F(4,294) = 0.66, p = .84$, see Table 7.2). We further explored whether topics and ranking impacted search diligence. However, two MANOVAS revealed neither evidence for topical differences ($F(5,293) = 0.98, p = .47$) nor for ranking effects ($F(2,296) = 0.74, p = .65$). Answering RQ2, not finding evidence for H2a-d, we could

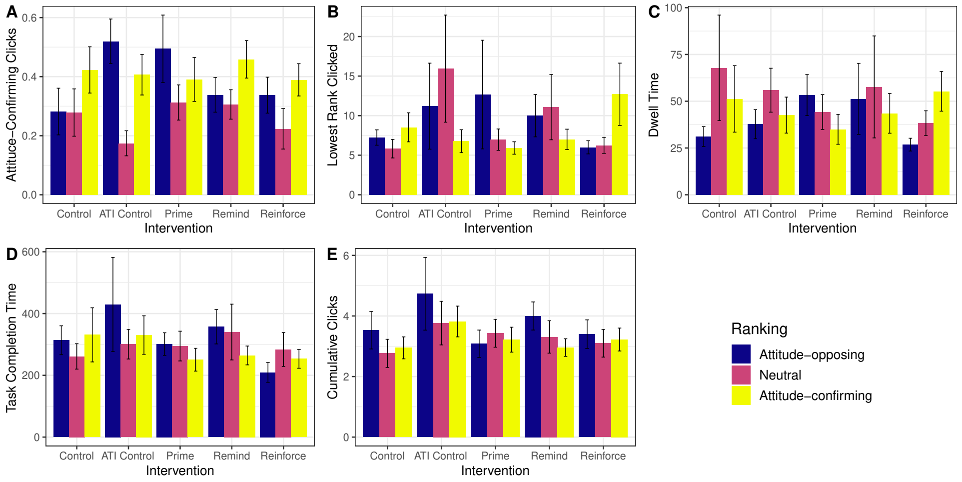


Figure 7.4: **Search Behavior per Intervention and Ranking.** Mean (A) proportion of attitude-confirming clicks, (B) lowest rank clicked, (C) dwell time, (D) task completion time, and (E) number of cumulative clicks per intervention (control, ATI control, prime, remind, reinforce) and ranking condition with 95% confidence interval.

not corroborate that any of the boosting interventions impact search diligence.

The lack of evidence for effects on search behavior across all three boosting interventions renders RQ3, aimed at identifying differences between the effects of the interventions, obsolete.

7.5.2. EXPLORATORY ANALYSIS

We aim to gain insights into user behavioral patterns that can (i) complement and add nuance to our findings related to our hypothesis tests, and (ii) address our exploratory research question. Due to their exploratory rather than confirmatory nature, we do not set a significance threshold. Nonetheless, we report statistical test results, including p-values, to highlight facets in our data that warrant confirmatory testing via future research.

Impact of IH. We investigated whether our data revealed relations between the level of IH and search behavior. Recall that we measured the level of IH only for participants in the remind ($n = 67$) and reinforce ($n = 63$) conditions since the IH questionnaire was only part of these interventions. We did not observe correlations between participants' IH and their proportion of attitude-confirming clicks ($r = -.09, p = .29$) or any of the variables used to capture search diligence. This evinced an absence of patterns hinting at decreased confirmation bias or increased search diligence of searchers with high compared to low IH.

To extend the understanding of the role of IH beyond search behavior, we investigated relations to searchers' self-reported knowledge gain and attitude change. Our explorations did not reveal differences across intervention conditions, nor a correlation between IH and attitude change. We did, however, observe a weak positive correlation between participants' IH and their self-reported knowledge gain ($r = .25$, $p = .004$), where searchers with higher IH reported higher knowledge gain.

Rationales for Behavior. To gain insights into whether the boosting interventions or the searchers' level of IH affected how participants approached the search task, we explored the rationales that participants reported (rationale for behavior). Guided by *RQe*, we were specifically interested in rationales indicating IH, reliance on the search result ranking, or confirmation bias. Thus, an expert annotator employed a mixed inductive-deductive open coding approach to identify the five distinct themes described below and categorize the rationales accordingly. Subsequently, a second expert annotator categorized a subset of 50 rationales, showing good inter-rater agreement ($\kappa = 0.76$).

1. **Intellectual Humility.** Participants who reported that their behaviour was guided by indicators of intellectual humility such as a desire to gain knowledge (indicating awareness of the limits of their knowledge), see arguments for different viewpoints (indicating awareness of the fallibility of their beliefs), or the good reputation of the source. E.g., *The results I clicked on were both for and against a vegetarian diet. I chose so because I wanted to see both sides of an argument.*

2. **Ranking.** Participants who reported that they followed the ranking when selecting search results. E.g., *I always click on the ones that appear first because they are more relevant.*

3. **Confirmation Bias.** Participants who reported that they clicked on search results in line with their opinions. E.g., *It aligned with my own views.*

4. **Content and Form.** Participants who reported that the title, snippet, or presentation of the search result sparked their interest. E.g., *Usually if something in the intro paragraph looked appealing. I also love list articles.*

5. **Task/Unclear.** Participants who reported that they selected search results to complete the task or their rationale was unclear. E.g., *To complete the task.*

7

We evaluated the proportion of participants per intervention condition who reported each rationale (see Figure 7.5). Similar proportions of participants in the control and boosting conditions reported to have relied on the ranking. Noteworthy, a lower proportion of participants in the boosting than in the control conditions reported rationales categorized as indicating *intellectual humility*. Overall, merely nine participants reported rationales that indicate confirmation bias. We did not see differences between the levels of IH of participants who reported different rationales. Therefore, neither the boosting interventions nor the searchers' level of IH affected the propensity to approach the search task with IH-related intentions or to rely on the search result ranking.

We explored behavioral differences across rationales (see Table 7.3). When contrasting confirmation bias and search diligence between searchers who reported IH-related rationales and those relying on the ranking we observed that the mean

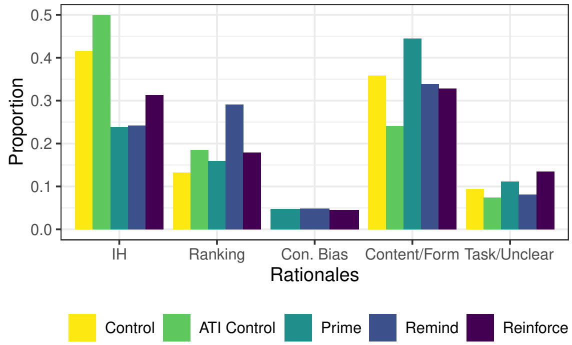


Figure 7.5: Proportion of participants in each intervention condition who reported a given rationale.

Table 7.3: **Confirmation Bias and Search Diligence per Rationale.** Means and standard errors yielded for each category of reported rationales behind search behavior.

	Attitude confirming clicks		Lowest rank clicked		Dwell time		Task completion time		Cumulative clicks	
	mean	*SE*	*mean*	*SE*	*mean*	*SE*	*mean*	*SE*	*mean*	*SE*
Intellectual Humility (n = 100)	0.3	0.03	9.5	1.38	54.3	6.63	348	33.6	3.7	0.28
Ranking (n = 57)	0.33	0.04	4.7	0.57	43.1	7.57	268	33.2	3.6	0.28
Confirmation Bias (n = 9)	0.58	0.12	6.6	1.21	30.6	10.4	318	85.8	3.8	0.85
Content and Form (n = 103)	0.38	0.03	10.4	1.58	42	4.4	279	16.7	3.2	0.2
Task/Unclear (n = 30)	0.45	0.07	8.7	2.08	39.2	16.5	244	44.0	2.7	0.3

proportion of attitude-confirming clicks was similar, while the mean values for lowest rank clicked, dwell time, and task completion times were higher for participants who reported IH-related rationales (see Table 7.3).

Reflections on the Search Task. To gauge if searchers' reflections on the search task are affected by the boosting interventions and related to search behavior, we examined the self-reported levels of mental demand, temporal demand, performance, effort, and frustration captured with the NASA-TLX questionnaire. We saw no differences in these measures between the control and boosting conditions, nor any correlations with confirmation bias and search diligence. We further explored whether the level of perceived task load varied for participants with different levels of IH. For the 130 participants for whom we captured IH, our exploration indicated weak correlations to mental demand ($r = -.18, p = .042$) and temporal demand ($r = -.21, p = .014$). Further, we observed moderate correlations to performance ($r = .4, p < .001$) and frustration ($r = -.44, p < .001$). Individuals with high compared to low IH reported lower mental and temporal demands, higher performance, and lower frustration. These relations suggest that high IH searchers might approach searching on debated topics with greater ease.

7.6. DISCUSSION

With this study, we investigated whether the potential of IH would indeed translate into unbiased and diligent search behavior among opinionated individuals seeking information on debated topics. For that, we compared the search behavior of participants exposed to one of three IH-boosting interventions with that of participants exposed to one of two control conditions. We also considered the role of IH during the broader search process by investigating measured IH, attitude change, self-reported knowledge gain, searchers' reported rationales for their search behavior, and reflections on the task, alongside search behavior.

Effect of Interventions on Search Behavior. We did not observe differences in searchers' confirmation bias and search diligence between the control and boosting conditions, yielding negative responses to RQ1 and RQ2, and rendering RQ3 obsolete. Still, outcomes resulting from exploring how IH factors into the search process (Section 7.5.2) allowed us to make some inferences that explain the absence of behavioral differences between control and boosting interventions. Furthermore, they point to alternative approaches for harnessing the power of IH for better search on debated topics.

The role of IH. We did not note direct links between the searchers' IH and their search behavior, which was unanticipated, given reports on prior research (see [40, 182, 268]). Searchers' reflections on the search task and their knowledge gain, however, suggest that searchers with high IH might approach searching on debated topics with greater ease and perceive to gain more knowledge than those with low IH. Looking at the rationales that searchers reported for their behavior, we saw that independently of the intervention and their level of IH, individuals approached the search task differently. For instance, some rely on the ranking, while others are driven by the desire to learn about diverse viewpoints. We infer that alternative factors that we did not consider in our study shape searchers' intentions as they approach the task. We explored the search behavior of participants who reported different rationales and observed that searchers who approached the search in an IH-driven way were more inclined to exhibit search diligence than those who reported having relied on the ranking.

Despite not finding effects from the boosting interventions on search behavior, we derive from these findings that IH and particularly IH-related search intentions seem to be relevant components in the pursuit of empowering opinionated individuals to fruitfully and with ease search for information on debated topics. As for why the interventions did not affect search behavior, our exploratory observations lead us to contemplate the following options:

1. **Familiarity of Search Environment:** To date, IH-boosts have been predominantly evaluated in terms of their effects on self-reported IH and reflection tasks [181, 268, 269], with less focus on their influence on practical behavior in a familiar information context. In contrast, we investigated the effect of the boosts in practice, on interactions with an interface designed to resemble widely recognized search interfaces. The familiar search environment might impede behavioral change, diminishing the effects of IH boosting

interventions administered prior to the search task, and potentially even of high IH as a general trait, by causing individuals to resort to their default search behavior (e.g., relying on the ranking). Resorting to default behavior in a familiar search environment resembles the phenomenon of *functional fixedness*, wherein individuals experience constraints to use a tool in unfamiliar ways [4, 79]. When exploring the reported rationales of search behavior, we observed that a high proportion of participants in both boosting and control conditions said that their behavior was driven by the ranking on the SERP. As suggested by Smith and Rieh [325], this indicates that searchers have learned to rely on the search system to compare, evaluate, and differentiate sources on their behalf. If interventions that boost IH or other cognitive skills do not cause behavioral change in strongly familiar and relied-on web environments, this would raise doubts about their general usefulness and emphasizes the importance of carefully assessing the effects of boosting interventions on behavior in the targeted web environments. Although boosting interventions are supposed to remain effective even after they were presented to the searcher [132, 210], future research should explore whether potential effects of familiarity and functional fixedness could be overcome with interventions that are more directly integrated into the search process rather than administered prior to it.

2. **Strong Attitudes:** Unlike prior research [40, 104, 182, 268], we did not find evidence for correlations between the level of IH and less biased or more diligent information seeking behavior. This could be attributed to the study's specific emphasis on strong attitudes. For instance, Krumrei-Mancuso and Newman [183] noted that individuals are less inclined to display IH when they interact with a topic for which they hold a strong attitude and their values feel threatened. However, in our pre-study, we observed boosted self-reported IH, even though we tailored the questionnaires to focus on a topic on which participants reported having a strong attitude. Yet, achieving behavioral change in practice is presumably more complicated than boosting self-reported reflections, and the strong attitudes may have acted as barriers, impeding any effect on search behavior. Future research should investigate how attitude strength and more nuanced attitude features, such as *attitude certainty* [345], or *attitude importance* [140] moderate the effects of various interventions to support unbiased and diligent search behavior.

Mitigating Confirmation Bias. Although the interventions did not noticeably affect search behavior, we made an unexpected yet intriguing discovery regarding a factor that did influence it: We observed that when a *neutral* search result was displayed on the top rank, participants exhibited lower confirmation bias than when an attitude-confirming or attitude-opposing search result was displayed on the same rank ($f = 0.21$). Further, none of the participants who saw a neutral search result on the first rank reported rationales related to confirmation bias for their search behavior. If this effect can be replicated in a follow-up study, displaying a neutral search result on the top rank during searches on debated topics could be one simple and practical approach to mitigate confirmation bias.

7.6.1. IMPLICATIONS AND FUTURE WORK.

Our exploration of searchers' reflections on the search task that we captured with the NASA-TLX indicate that individuals who exert more effort perceive less frustration and a sense of better performance upon completion of the search process, suggesting a fruitful and satisfying search experience. This reassures us that delving into ways to promote unbiased and diligent search behaviour on debated topics ultimately benefits not only searchers' informedness but also their search experience. However, the interventions we considered in this study boosted self-reported IH, but did not foster unbiased and diligent search behavior in practice.

Reconsidering the question of how to empower individuals to overcome the challenges associated with web search on debated topics in light of our newly gained understanding of the shortcomings of the tested interventions and the role of IH in the broader search process, we conclude that a standalone solution likely does not exist. Instead, we need a combination of measures that address different challenges associated with searching on debated topics. For example, our explorations indicated that it is not necessarily individuals with high levels of IH but those who approach the search task with IH-related motivations who tend to exhibit more diligent search behavior. If future research confirms this relation between search IH-related search intentions and diligent search behavior, interventions should more directly **motivate IH-related search intentions**. Further, we learned that the boosting interventions did not modify searchers' reliance on the search system and ranking. Hence, there is a need for strategies to **support appropriate reliance** on the search system and ranking in familiar search environments and overcome effects of functional fixedness, for instance by more directly integrating interventions into the search process or enriching the knowledge-context in SERPs with epistemic cues [210, 325]. Moreover, search environments that individuals tend to over-rely on could be redesigned to earn that reliance. For instance, we observed that displaying a search result with a neutral stance on the top rank might be a practical approach to mitigate confirmation bias. Approaches to re-rank search results to increase the viewpoint diversity among highly-ranked search results, as suggested by Draws et al. [72], should be considered as a fundamental part of the solution.

As for efforts to empower individuals online more generally, our findings illustrate that the effects of interventions on behavior need to be carefully investigated in the target environment. This supports the cautionary stance by Freiling et al. [91] who warn against the hasty deployment of interventions to guide online information behavior while disregarding the complexity of the problems they aim to overcome and the broader ethical implications of the interventions. That said, we should keep in mind that search systems without interventions are far from being neutral gateways to information. On the contrary, search systems act as algorithmic curators [342, 372] that are predominantly under the control of private industry [95] and thus designed to prioritize commercial interests. This is showcased by persuasive and manipulative choice architectures, such as featuring sponsored content among the top-ranked search results [178, 180, 398].

From this study, there are several avenues of future research to embark on. First, the effect of more directly integrating interventions into the search process, combining different measures that boost IH, motivate IH-driven search, and promote transparency for appropriate reliance on the ranking, e.g., by applying epistemic cues such as stance labels deserves thorough investigation. To pinpoint interventions that motivate IH-driven search, future research should strive to uncover factors that shape searchers' intentions as they approach search on debated topics. Moreover, the preliminary finding on the impact of placing a search result with a neutral stance on the top rank as a practical approach to mitigate confirmation bias suggests the need for a more focused study design to delve deeper into this phenomenon. Lastly, in light of the increasing significance of passive information exposure in contrast to active information seeking, as highlighted in recent work by Hassoun et al. [125], the role of IH in various information settings that extend beyond web search warrants investigation.

7.6.2. Limitations

An in-depth user study such as the one we undertook is not without limitations. For data gathering, we used a mock search interface that mimics conventional search systems. During the study, participants could issue multiple queries and access several SERPs; however, all SERP results were derived from a preselected set of viewpoint-annotated search results and ranked to conform with our ranking templates of alternating viewpoints rather than relevance to the query. The ranking templates—employed to control for ranking effects on participants' search behavior—led to interactions with diverse viewpoints, regardless of whether individuals relied on the ranking or actively sought to engage with diverse viewpoints and consequently may have contributed to overall low confirmation bias across conditions. Future work should consider the role of IH in scenarios that fully reflect the complexities of real-world web search, including those when searchers are exposed to SERPs featuring viewpoint-biased rankings where most highly ranked results align with a single viewpoint.

Given the intent of this study (investigating if interventions boosting self-reported IH could affect search behavior), we captured the level of IH solely for participants who were part of the interventions involving the IH questionnaire ($n = 130$). In their case, we noted relatively high levels of IH, indicating a somewhat skewed sample. We were surprised by the large percentage of participants in the control group, as opposed to the boosting conditions, who reported IH-related rationales. This could indicate an unequal distribution of participants with different levels of IH across the five intervention conditions, which we could not control for, as we lacked information on participants' levels of IH in the control and prime conditions. Future studies should measure the level of IH of all participants and consider recruitment strategies aimed at achieving a distribution of participants with different levels of IH that is more closely aligned with that of the general population.

To assess whether our data could have been negatively impacted by cognitive biases provoked by the task design of the crowdsourced user study, we applied the Cognitive Biases Checklist introduced by Draws et al. [74]. Similar to most studies

7

relying on crowdworkers, *self-interest bias* could have affected the search interaction data–participants may have invested minimal effort to complete the task and receive the reward [171], and thus deviated from their usual search behavior. However, only data from participants who passed the attention checks, included to counter this bias, was considered for analysis.

7.7. CONCLUSIONS

With this user study, we investigated whether IH-boosting interventions could contribute towards empowering opinionated searchers to overcome the challenges associated with search on debated topics by fostering unbiased and diligent search habits. For that, we investigated the effect of three boosting interventions on search behavior in a familiar search environment, as well as the role of IH in the broader search process. We found that the interventions that boost self-reported IH did not result in searchers adopting unbiased and diligent search behavior in practice. Our exploratory findings indicate that both IH and IH-related search intentions are nonetheless relevant elements for cultivating unbiased and diligent search behavior, as well as a fruitful and satisfying search experience.

In light of our findings, we advocate for comprehensive interventions to not only boost IH but also motivate IH-related search intentions and support appropriate reliance on the search system and ranking in familiar search environments, for instance by being more directly integrated into the search process.v Moreover, outcomes from our exploration emphasize the importance of thoroughly investigating the effects of interventions that aim at empowering individuals online in practice, with a focus on their impact on behavior within the target environment, rather than solely on self-reflection or on performance in simulated tasks.

IV

Towards Supporting Responsible Opinion Formation

Assessing our combined insights, we recognize that supporting responsible opinion formation in web search on debated topics is a complicated undertaking of which we are merely at the starting point. However, we also see an urgent need to identify solutions that genuinely mitigate rather than unexpectedly increase harm [91, 243]. In Part IV we hence target the following research question:

> **$\textbf{RQ}_{\textbf{IV}}$**: What are challenges and research opportunities towards supporting web search on debated topics, promoting responsible opinion formation?

In Chapter 8 we provide a foundation for future research efforts *towards enabling responsible opinion formation* in web search on debated topics. For that, we built on perspectives from digital humanism and, through an extensive interdisciplinary literature review, outlined challenges and research opportunities inherent to the searcher, the search system, and their interplay.

8

Responsible Opinion Formation on Debated Topics in Web Search

This chapter is based on a full conference paper: **Alisa Rieger**, Tim Draws, Nicolas Mattis, David Maxwell, David Elsweiler, Ujwal Gadiraju, Dana McKay, Alessandro Bozzon, and Maria Soledad Pera. "Responsible Opinion Formation on Debated Topics in Web Search". In: *Advances in Information Retrieval*. Ed. by Nazli Goharian, Nicola Tonellotto, Yulan He, Aldo Lipani, Graham McDonald, Craig Macdonald, and Iadh Ounis. Cham: Springer Nature Switzerland, 2024, pp. 437–465. ISBN: 978-3-031-56066-8. DOI: 10.1007/978-3-031-56066-8_32. Alisa Rieger and Tim Draws contributed equally to that paper.

8.1. Introduction

Conventional search engines fall short of aiding complex, consequential information needs [105, 226, 320, 325], prompting the question *how web search can support information seeking on debated topics*. By that, we do not mean guiding searchers toward a particular view or ideology but instead assisting and empowering them in actively and thoroughly engaging with diverse viewpoints; critically evaluating information to **form opinions responsibly** [177, 260]. Although users may intend to expose themselves to diverse viewpoints when searching for debated topics [4, 219], responsible opinion formation can be impeded by factors like over-relying on the system to provide accurate and reliable resources [325]. Engaging with information on debated topics is naturally demanding and can trigger emotionally charged behavior, as it has the potential to challenge the searcher's core beliefs and values [140, 267]. Thus, search on debated topics inherently requires cognitive effort, particularly to overcome *biases* that can occur during the search process. Such biases may emerge from the user (e.g. cognitive biases) [137, 376, 383], the search engine (e.g., data, relevance criteria, and algorithmic ranking biases) [72, 103, 356], or the interaction between them (e.g., presentation, over-reliance, and contextual biases) [18, 22, 325]. These considerations highlight the complex, mutually evolving interplay of the searcher and the search engine (see Figure 8.1), as illustrated in representations of the search process such as the *Information-Seeking and Retrieval Model* [146, Chapter 6].

As search engines are widely used, they can and should be platforms to explore debated topics in all their nuances. The information retrieval (IR) community has dedicated efforts to comprehending the evolving needs of searchers and society and developing technology to support them [250, 320]. Given the role search engines play in opinion formation—a search intent they were not explicitly designed for—the importance of advancing the understanding of the associated challenges, as well as the development of system functions that foster responsible opinion formation, becomes apparent. Although IR research has already explored and experimented with *fairness* [14, 94, 393], *diversity* [2, 76, 312], *argument retrieval* [38, 78, 259, 272, 366], and user interface adaptations [52, 158, 212, 386], whether and how web search engines should cater to users' opinion formation and deal with debated topics remains largely unanswered. Resonating with the ideals for future technological development of *digital humanism*, web search should be shaped following individual and societal values and needs instead of letting web search shape individuals and society [371]. To do so, it is essential to recognize opinion formation on debated topics as a distinct search intent, characterized by (1) the heightened risk of searcher and search engine biases and (2) its consequential nature on individuals and society at large, and warranting dedicated research efforts. The IR community is uniquely positioned to spearhead interdisciplinary efforts to advance such socio-technical research endeavors.

In this chapter, we examine the role of web search engines in users' opinion formation, delineating the distinct characteristics of web search on debated topics through an extensive review of interdisciplinary literature. We illuminate the challenges inherent to the searcher (Section 8.3), the search engine (Section 8.4), and

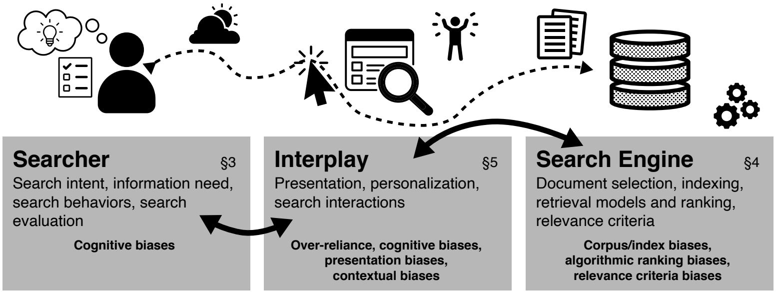


Figure 8.1: Search on debated topics. Biases hindering interactions to gain well-rounded knowledge can emerge from the searcher, the search engine, and their interplay. Ultimately, search on debated topics can shape cognitive processes (e.g., attitude change) and concrete actions (e.g., voting in an election).

their interplay (Section 8.5) and outline a research agenda (Section 8.6) encompassing methodological considerations, high-level challenges, and initial research questions towards responsible opinion formation through web search.

8.2. Digital Humanism and Responsible Opinion Formation

Digital Humanism advocates for reflecting on the relationship between humans and technology. Fostering human-centered design, it prioritizes better lives and societal progress over mere economic growth [371]. Designing technology to embody these ideals is not a linear process as technology and humans co-evolve, mutually shaping one another in an intricately intertwined manner [248, 380].

Web search is one of the primary information gateways, impacting searchers' knowledge, choices, and actions [48]. Searchers have cultivated a sense of trust that makes them rely on the system's evaluation and differentiation of resources on their behalf [325]. Yet, search engines are not subject to regulations for content quality and diversity necessary for an informed citizenry, unlike the standards applied for responsible reporting within traditional media outlets [130]. Opaque relevance and ranking criteria are far from value-neutral but function as algorithmic curators that serve a goal, e.g., user satisfaction and profit generation [48, 246, 372]. Given the profound impact of web search, recent work has called for revisiting relevance criteria and search system design to better align with the needs and values of individuals and democratic societies [44, 105, 320]. However, it is non-trivial to balance values that might be in tension with each other [48]. These tensions are particularly evident for search on debated topics, where relevance to user needs might not be aligned with relevance to democratic values, necessitating a critical evaluation of value trade-offs.

Forming opinions responsibly involves gathering evidence and critically assessing it [177, 260]. In the context of web search, this translates to searchers actively and thoroughly engaging with search results encompassing **diverse viewpoints**. Yet, this

8

is not the norm as SERPs often lack viewpoint diversity [72], and searchers tend to primarily interact with information that aligns with their own viewpoints [302, 323, 360, 372].

Viewpoint diversity in people's exposure to information concerning debated topics represents a long-standing research topic in the communication sciences [21, 37, 84, 207, 381]. Different democratic notions of viewpoint diversity can be applied depending on the objectives of a system [128]. Which particular notion of viewpoint diversity is appropriate in an opinion formation-related search scenario, however, might depend on both the topic and the user [128, 363, 364]. For instance, one could argue that viewpoint diversity is vital for unresolved issues but that web search engines should represent topics with a solid scientific basis in a more one-sided fashion. While it may seem obvious that scientifically answerable topics should be presented as such, previous research has shown that exposing strongly opinionated users to nothing but opposing viewpoints can result in a backlash effect; where they become more entrenched in their beliefs [249]. This can increase polarization by leading users to shift their attention away from mainstream toward more niche information sources [249]. Similarly, increased diversity can also lead to false perceptions of existing evidence, e.g., balancing climate change believers and deniers can create a false image of an open debate that may be worse than an approach that accounts for different weights of evidence [64]. The desirable degree of viewpoint diversity may thus not always be either the minimum or maximum [23] and can depend on the topic and individual user characteristics [216, 239].

IR research has largely used binary (e.g., democrat/republican) or ternary taxonomies (e.g., against/neutral/in favor) [98, 266, 280, 388] as *viewpoint representations* for search results. Recent work, however, has shown that such labels unnecessarily reduce complex viewpoints to generic categories, which limits the insight gained in research using them [68]. Researchers have added more nuance to such labels by using ordinal scales [74, 298], continuous scales [186, 187], multi-categorical perspectives [54], or building on outcomes from communication sciences [20, 37] to yield a two-dimensional viewpoint label that includes a nuanced notion of *stance* (e.g., strongly supporting) and *logics of evaluation* (i.e., representing the reasons underlying a stance, e.g., supporting zoos because of their animal conservation efforts) [68]. Despite these advancements, there is a need to analyze existing viewpoint representation frameworks for comprehensibility, practical applicability, and meaningfulness for users and practitioners.

8.3. The Searcher

The searcher (*information seeker*) turns to a search engine to execute a search intent motivated by an underlying *information need*. This develops from a perceived problem, a knowledge gap, an internal inconsistency related to their understanding, or some conflict of evidence [28]. Once the searcher enters a *query* into the system, their interaction with the system begins (Section 8.5). Such interactions include evaluating the information encountered in search results and can affect searchers' knowledge and attitude towards the search topic [82, 173].

Research on how users search the web for debated topics [137], or how they form opinions in non-biased scenarios [99, 219] is in its infancy. Progress depends on conducting user studies into behavioral patterns as users search for debated topics (e.g., queries used [4], if they engage with counter-attitudinal viewpoints, or when they stop searching) and searchers' preferences (e.g., whether users prefer diverse or filtered viewpoints [167]). Also crucial are methods to correctly interpret user behavior, e.g., clicks on search results are often seen as a proxy for engagement [74, 298] but users may engage with them in a variety of ways that can be just as meaningful for opinion formation [176]. Researchers should investigate how to support users' reflections on their search processes and outcomes (e.g., awareness of their biases and knowledge level) and investigate long-term opinion formation (e.g., changes in search behavior and opinions over time).

Cognitive Biases. To reduce the cognitive demands of processing information on debated topics, searchers frequently (and subconsciously) employ shortcuts, which can introduce cognitive biases [18, 99, 351]. *Confirmation bias*, searchers' tendency to prioritize information that confirms prior attitudes [244, 360, 372], can prevent engagement with diverse viewpoints during search on debated topics. This bias has been observed at various stages of the search process, e.g., query formulation [137], and search result selection [253, 360, 383]. Other studies have noted searchers' inclination to engage with positive (i.e., query-affirming) [376] and mainstream content [99]. Triggered by the search result presentation, other cognitive biases that hinder diligent search behavior can arise (Section 8.5). Identifying how to facilitate search in this context requires a thorough understanding of factors affecting searchers' intentions, behavior, vulnerability to biases, and evaluation of the encountered information. It also requires approaches to support and empower searchers for unbiased and diligent search behavior.

Context. The vulnerability to biased search behavior is contingent upon the searcher's context. For instance, when searching purposelessly, as opposed to specifically looking for information on a particular debated topic, searchers' vulnerability to cognitive biases increases [383]. Stressful conditions (e.g., time pressure) may strengthen the influence of cognitive biases [264, 316]. This calls for investigating how the searcher's context influences search behavior and the vulnerability to cognitive biases when engaging with debated topics; also how to create search environments that foster unbiased and diligent search behavior and reduce contextual conditions leading to high vulnerability to biases.

User Characteristics. Search behavior, susceptibility to cognitive biases, and reaction to elements of the user interface are affected by *situational* and *stable* user characteristics [353]. Situational factors include attitude strength and certainty [172, 358] and involvement with and prior knowledge of the topic [205, 237, 377]. Stable factors that affect engagement with debated topics include searchers' *need for cognition* (i.e., an individual's tendency to organize their experience meaningfully) [47, 261, 348], *receptiveness to opposing views* (i.e., willingness to impartially access and evaluate opposing views) [229], and *intellectual humility* (i.e., an individual's tendency to recognize the fallibility of their beliefs and the limits of their knowledge) [40, 67, 104, 182, 195, 267]. Open research directions include advancing the understanding of how different user characteristics affect search on debated topics throughout the search process, from search intent to search evaluation, and if concepts such as searchers' *moral values* [204, 291] play a role. Researchers should also investigate how efforts to support unbiased and diligent search behavior may require adaptation to cater to the diverse needs of searchers with distinct characteristics.

Vulnerable Groups. It is crucial to study and accommodate vulnerable user groups such as children, elderly people, or neurodivergent users in search for opinion formation. These users have certain characteristics (e.g., fewer cognitive resources or low technological literacy) that may make them more vulnerable to viewpoint biases and less likely to enact responsible opinion formation [165, 189, 208, 228]. For instance, children are less likely to judge or explore search results [189] and are more susceptible to opinion formation through misinfomation [208]. Elderly users similarly have increased tendencies toward sharing and interacting with fake news [111, 155]. Research is needed to identify who those vulnerable groups are specifically, what particular factors make them vulnerable, and how web search engines can support these users in their opinion formation.

Boosting Searchers' Competencies. Boosting interventions are effective in fostering web literacy skills, such as resilience to misinformation [200, 304], detecting micro-targeting [209], and improving privacy behavior [252]. These interventions, which promote individuals' cognitive or motivational competencies [132, 178, 210], contain a learning component and thus could remain effective even after the intervention. The specific challenges posed by web search on debated topics might require an expansion of traditional web and information literacy constructs [114], for instance, by incorporating intellectual virtues [105]. Although boosting interventions that target such virtues have been suggested [295], their effect on search behavior and opinion formation is not fully understood.

8

8.4. The Search Engine

Contemporary search engines provide a means of sifting through large volumes of information to find the proverbial *needle in the haystack*. Key to search engines are three inputs: (i) a **document index**, a data structure representing a collection of documents (or corpus, typically a *crawled* [166] collection of web documents for web search engines); (ii) a **retrieval model**, that is responsible for identifying and scoring (and ranking) documents that are deemed *relevant* to what is being searched for, based on a series of *relevance criteria* (e.g., [152, 330, 387]); and (iii) a **query**, a construct of an *information need* as provided by the searcher, typically formulated as a series of tokens, e.g., `'should zoos exist'`. Search engines—as with other systems—are not immune from biases [236]. Indeed, the design of the retrieval model can raise several areas in which biases can (and do) arise, such as leading to *undue emphasis* on particular perspectives [187].

Corpus/Index Biases. Search results can only list documents that are included in a web search engine's index. With commercial web search engine crawlers indexing huge swathes of the *World Wide Web*, the population of content creators who generate the documents in this collection is unlikely to represent the global human population [103], and follows a highly unequal distribution concerning the number of documents generated per content creator [10, 13, 357]. Such collections may thus include a *creation bias,* i.e., they do not contain balanced or society-representative viewpoint distributions on all debated topics [254, 327]. Moreover, the way in which a retrieval system indexes documents can affect the distribution of available documents. An *indexing bias*—whereby the search engine is programmed to systematically ignore particular documents—may further skew the data that the retrieval system can process [265, 356, 359].

Algorithmic Ranking Biases. Search engines may (unintentionally) exacerbate viewpoint biases in the indexed corpus through algorithmically-biased relevance criteria [94, 255, 256]. *Ranking biases* may cause documents that express certain viewpoints to rank higher than others, and therefore receive more attention from searchers (Section 8.5). This can occur when search result rankings solely focus on relevance criteria that optimize for maximizing searchers' satisfaction [354].

Relevance Criteria Bias. Determining the relevance of a search result is central to search engines. With debated topics, the relevance criteria employed by conventional search engines—which mostly target user satisfaction to maximize profit and efficiency [354]—may prove inadequate. Disregarding relevance to the unbiased knowledge gain of the searcher—as well as relevance to society and public welfare—can impede searchers from gaining a comprehensive understanding of a debated topic and its various arguments [105, 115, 130, 330]. Prior work has found viewpoint biases in highly-ranked search results concerning health information [372, 375], politics [279], and other debated topics [72].

Research and practical applications require automatic viewpoint classification methods to evaluate and foster viewpoint diversity. This primarily concerns the development of bias metrics and diversification algorithms.

Viewpoint Detection. Applications for search on debated topics need efficient and reliable methods to assign viewpoint labels to documents, e.g., measuring or mitigating search result viewpoint biases in real time. Recent research has seen the emergence of *Natural Language Processing* (NLP) tasks like *stance detection* [16, 118, 231, 232, 329, 368] and *argument mining* [43, 191, 192, 203, 245, 331], which aim to automatically detect different viewpoint components in text. Other works have used unsupervised topic models [341, 346, 397] or hybrid approaches (i.e., automatic methods combined with crowdsourcing) [19] to overcome the limitations of supervised stance detection models. However, practitioners will ultimately need fully automatic methods to classify search results into broad viewpoint representations. *Large Language Models* (LLMs) have recently shown promise in this area, still further work is needed. Researchers should build on the existing efforts in stance detection, argument mining, and argument retrieval [3, 38] to develop such advanced methods.

Viewpoint Bias Assessment. Assessing viewpoint bias requires metrics that accommodate the chosen ethical notion of viewpoint diversity and viewpoint representation. Current rank-aware viewpoint bias metrics applicable to search results consider categorical stance labels (e.g., against/neutral/in favor) [363, 385], continuous stance labels (e.g., ranging from -strongly opposing to strongly supporting) [187], or multi-dimensional viewpoint labels (i.e., stance and logic of evaluation [72]. Thus far, viewpoint biases in search results are primarily assessed as a deviation from viewpoint balance [72, 75, 82], deviation from the overall distribution across ranks [94, 187], or the presence of scientifically false information [266, 375]. Yet, it is unclear what metric may best apply in what scenario, how metrics compare, and what intuitive degrees of viewpoint bias different metric scores suggest. Existing metrics do not distinguish among data, algorithm, or presentation bias, and there is no guideline as to what specific *discount factor* to apply for rank-awareness [313]. There is a need to develop comprehensive viewpoint bias metrics, (simulation) studies to compare metrics, interpretation guidelines (i.e., including metric thresholds where viewpoint biases may become problematic), and best practices for using those metrics.

Viewpoint Diversification. Earlier work has diversified search results for more general user intents [2, 142, 160, 312, 395], and even made first steps to manually or automatically diversify viewpoints [72, 83]. While some of these works have considered advanced viewpoint labels [72], how to diversify search results for different diversity notions or viewpoint representations, and how to dynamically adapt diversification algorithms to searcher needs (e.g., due to changes in search topic or user context) remains to be determined. Researchers could further explore solutions for data, algorithmic, and presentation biases individually and develop pipelines that increase diversity at each level.

8.5. The Searcher and Search Engine Interplay

Search engines present the SERP to the searcher, featuring search results that may be personalized, taking into account several contextual factors, such as previous search interactions [174, 336]. Searchers interact with the SERP, for instance by querying, scanning the results, and clicking on selected items to access the web page. Substantial challenges associated with searching on debated topics emerge from the intricate interplay of the searcher and the search engine.

Over-Reliance and Cognitive Biases. Searchers rely on search engines and assume that highly-ranked search results are relevant and accurate [99, 325] — a notion that may be explained with the perceived quality of top-ranked results (e.g. see work on the related context of news selection [108]), or as a response to information overload. Indeed, prior work shows that when the amount of available information exceeds one's processing capacities, searchers tend to be more selective and prone to cognitive biases [332]. For complex tasks, this reliance may impede searchers from expending the needed cognitive effort, thus turning into over-reliance [325]. Opaque relevance criteria further hinder searchers' ability to assess information completeness [226]. Reliance on the search engine is exemplified in searchers' *position bias* (i.e., users typically tend to pay much more attention to search results at higher ranks [150, 257]) as well as the *search engine manipulation effect* [33, 82, 266], where users tend to change their attitudes following viewpoint biases in search results. So far, little prior work has explored what gives rise to phenomena such as the search engine manipulation effect [74]. Effects emerging from the interplay between the searcher and search engine might also be related to additional cognitive biases, such as the *availability bias* (i.e., overestimate the prevalence of information that is easily accessible) [18], or *anchoring bias* (i.e., the top-ranked search result may color the searcher's attitude) [18, 247, 376]. Such phenomena typically occur without users' awareness [99] and are unlikely what users aim for when they search the web for debated topics. Moreover, as web search results get increasingly augmented or replaced by highly pleasing and personalized answers from artificial intelligence chat systems (e.g., *ChatGPT*) that require exerting even less cognitive effort when searching, over-reliance and cognitive biases among users may become even more prevalent.

Presentation biases. Search results are typically presented as ranked lists (i.e., split into pages of ten search results each; although other presentation formats have been proposed [154]). Each result is displayed with a *title*, a *snippet* (i.e., a brief excerpt from the document text), and the relevant *URL*. Common web search engines often display additional information such as *entity cards* [39], direct answers [31], or suggestions for alternative queries [220]. These different factors provide ample room for presentation biases in search results [22, 24, 390]. Viewpoint-related presentation biases could occur due to a more prominent presentation of particular viewpoints, e.g., by more favorable snippets [33, 34] or representation in entity cards [212]. Moreover, the impact of presentation biases could be largely hidden as users often engage with search results without clicking on them (e.g., only reading the titles and snippets) [176].

Context. Contextual factors emerging from the searcher-system interplay may aggravate biases [145]. For instance, search result rankings may be affected by users' prior searches, preferences, or location [251, 392], viewpoint biases in earlier interactions may lead to biased follow-up search queries [4], and presentation biases may depend on the device that users employ [168].

The biases and artifacts arising from the mutually evolved interplay between searchers and search engines can obstruct fruitful searches that facilitate responsible opinion formation. Thus, there is a need to disentangle and understand this convoluted interplay and design search interfaces that facilitate and motivate thorough engagement with diverse viewpoints.

Exposure and Interaction. The search results users are exposed to (and subsequently interact with) can strongly influence users' opinions [7, 33, 82, 266]. *How* users interact with search results plays an important role here: even when exposed to viewpoint-biased search results on social and political information, search behavior is still characterized by searcher-rooted interaction bias, with searchers prioritizing search results that align with their beliefs [302, 323]. While searchers may somewhat defy the impact of exposure effects, they could still lead to more subtle and enduring consequences over time [302]. These observations stress the need for deeper insights into the dynamics of exposure and interaction biases. Considering that viewpoint changes often begin with information encounters on social media [125, 219], researchers should moreover explore the relation of exposure and interaction effects across different information settings.

Interfaces. Interface modifications can support unbiased and diligent search behavior, e.g., presenting search results in alternate formats [158], providing information about the search topic or the ranking [212, 386], visualizing viewpoints and biases in search results [52, 83, 382], suggesting alternative queries [270], or highlighting documents with diverse viewpoints [56, 388]. Also promising are behavioral interventions to support unbiased search interactions (e.g. warning labels) [83, 261, 297, 298]. Researchers should investigate how different viewpoint representations, notions of viewpoint diversity, and additional features, e.g., search result explanations, affect searchers [70, 286, 382]. Interventions that can be customized by the searcher (i.e., *self-nudging* [287]) have worked in the news context [27, 120, 355] and merit investigation in the realm of web search. As users sparingly utilize customization features and adhere to default configurations [338, 355], research is needed to identify user-friendly options and optimize default settings. Increasing search engine *transparency* (e.g., by explaining what factors influenced the ranking or providing meta-information for search results) as a means to raise awareness of system biases and foster appropriate reliance should be investigated. This could boost searchers' technological and information literacy [131, 325]. Still, providing meaningful explanations poses several challenges, including decisions regarding the level of detail and presentation [77].

Personalization. Users have diverse characteristics, tendencies, and pre-search opinions [33, 74]. This raises the question of whether degrees of viewpoint diversity or presentation formats (e.g., stance labels) should be adapted to different searchers [290, 298]. Personalization with regards to searchers' opinions, cognitive biases, moral values, and other relevant constructs would require methods to automatically predict these psychometric variables [218]. However, such endeavors would also raise substantial privacy concerns [344]. Whether and how to customize search results and the interface based on factors like user characteristics, past behavior, and the specific topic remains an open question that warrants ethical and research discussion. This may also affect general personalization efforts by web search engines [174, 274, 336].

8.6. Research Agenda

The intricate dynamics among the searcher, the search engine, and their interplay (Section 8.3- 8.5) call for reflecting on research methods and broader research challenges. We outline some of these considerations and challenges, along with research questions to guide efforts on web search on debated topics.

Data Collection and Public Datasets. Developing and evaluating methods to assign viewpoint labels or foster viewpoint diversity in search results, and user studies on search behavior require high-quality, human-annotated ground truth datasets with search results and viewpoint labels. Creating such datasets is not easy: recent research has shown that different worker characteristics and cognitive biases can reduce the quality of data annotations, especially in subjective tasks such as annotating viewpoints [71, 75, 80, 143]. More work is needed to identify best practices and publish openly available datasets with search results and comprehensive viewpoint labels for different debated topics.

User Studies. Evaluating perceptions of viewpoint representations and viewpoint diversity, understanding factors influencing searchers' behavior, and determining how to support unbiased and diligent search requires qualitative and quantitative studies. Winter and Butler [380] stress the value of ongoing dialogues between the technology developers and users, for responsible technology design. As we investigate issues concerning information access and societal well-being, it is crucial to comprehensively and longitudinally assess design choices and interventions in real-world settings, ensuring that they do mitigate harm rather than inadvertently exacerbating it [91]. Comprehending the impact of various factors on searchers and their behavior needs carefully designed, controlled studies with large sample sizes to grasp subtle differences [193]. Simultaneously, the uncertainty of the complex socio-technical dynamics, normative dimensions, and related risks might necessitate more exploratory research methods [315]. A promising new avenue in this regard that has recently gained traction in the communication sciences may be data donations. While they present legal, ethical, and technical challenges, data donations offer externally valid and highly granular insights by enabling researchers to retroactively analyze authentic search queries (e.g., from donated browser histories) [11, 36, 138].

Cultural Diversity. Different societies, countries, and cultures have vastly different ways of searching about and discussing debated topics [144, 188]. Contemporary academic research is almost exclusively conducted in English, and so is previous work related to search on debated topics. Yet, searchers across the globe may experience viewpoint biases and their undesired effects. It is, therefore, essential that future research considers web search on debated topics and all related challenges from a multi-lingual and multi-cultural perspective.

Misinformation. Balancing the dangers of exposing users to search results containing false claims with viewpoint diversity while preserving freedom of speech and avoiding (perceptions of) censorship is a particularly difficult issue that requires further investigation. Researchers and practitioners who work in the search for opinion formation space should be aware that misinformation may be particularly impactful here, and therefore closely monitor and leverage ongoing research efforts on misinformation detection and mitigation [88, 333, 394, 396].

Alternative Search Paradigms. In this chapter, we have focused on the traditional and dominant idea of search engines that present results as ranked lists. However, there are several alternative paradigms for which the retrieval process, result presentation, and user behavior diverge. Considering these differences becomes pivotal when designing interfaces that synthesize results from different resources into seemingly relevant and coherent written or spoken text [285, 288, 347]. Conversational interfaces are relatively more engaging than conventional web interfaces in various contexts [17, 112, 234, 281], including potential in supporting long-term memorability [282]. Notably, the pursuit of improving user engagement and experience can be orthogonal to supporting responsible opinion formation.

This dichotomy is perfectly captured by the well-established notions of '*seamless*' versus '*seamful*' design in human-computer interaction (HCI). While seamless design emphasizes clarity, simplicity, ease of use, and consistency to facilitate interaction with technologies, seamful design emphasizes configurability, user appropriation, and the revelation of complexity, ambiguity, or inconsistency [147]. There are several arguments in favor of creating *seamless* interactions with search systems to satisfy user information needs. However, such design choices may not adequately foster responsible opinion formation. Users may also turn towards LLM-based tools like *ChatGPT* [241], which may provide incomplete, misleading, or even inaccurate information due to model hallucinations. Natural language aids comprehension and offers opportunities to directly provide diverse viewpoints (i.e., serving as a *seamless* mode of interaction). However, Shah and Bender [320] warn that such interactions can hinder users' ability to identify incorrect or biased information and to actively explore different resources to construct a model of the knowledge space, building information literacy (i.e., facets that can be supported through *seamful* design). More research is urgently required to better understand whether and how responsible opinion formation can be supported in the context of such emerging search paradigms.

Malicious Intent. Thus far, we have assumed no malicious intention from any actor, i.e., framing biases and harmful effects as unintended byproducts of web search. Yet, malicious actors may use research findings and practical applications for their purposes, e.g., to steer public opinion or manipulate targeted individuals. This solicits methods to detect and safeguard against such actions. Researchers and practitioners need to discuss this possibility in their work.

RESEARCH QUESTIONS

The research opportunities and challenges discussed in this chapter may appear abundant and intimidating. To provide a more approachable starting point, we propose a set of research questions, which are by no means exhaustive.

Foundations: (i) What obligations should search engines bear concerning individual and societal well-being? (ii) Which values and principles should guide the system design process? (iii) What framework can comprehensively represent viewpoints on SERPs? (iv) Which notions of viewpoint diversity would benefit individuals and society? (v) Should the notion of viewpoint diversity be adjusted depending on the specific topic and searcher?

Searcher: (i) Which patterns of search behavior and searcher characteristics can be linked to knowledge gain and attitude change? (ii) Which traits affect searchers' vulnerability to ranking and cognitive biases? (iii) What user-centered interventions can empower unbiased and diligent search behavior?

Search Engine: (i) How should relevance criteria be adjusted for search on debated topics? (ii) What crowdsourcing, automatic, or hybrid methods can accurately and efficiently detect viewpoints expressed in search results? (iii) Which re-ranking strategies meaningfully increase viewpoint diversity?

Interplay: (i) What factors shape the interplay of search engine-rooted exposure biases and searcher-rooted interaction biases? (ii) What interface-centered interventions can empower unbiased and diligent search behavior? (iii) How can the interface be leveraged to enhance the transparency of relevance criteria to the searcher?

8.7. CONCLUSION

Drawing upon perspectives from digital humanism and an extensive body of interdisciplinary literature, we offer an in-depth analysis of the distinguishing characteristics and challenges associated with web search on debated topics. We outline a research agenda toward web search that fosters responsible opinion formation by focusing on the searcher, the search engine, and their complex interplay. While rooted in IR, advancements in this area demand a multi- and interdisciplinary approach with input from various domains, including philosophy, psychology, information science, and the communication sciences. With this chapter, we aspire to motivate researchers, practitioners, and policymakers across domains to engage in the collective effort of addressing the pressing socio-technical challenges and creating an enriching, unbiased, and trustworthy web search experience. Ultimately, the pursuit of such endeavors would benefit both individuals and society by promoting democratic values, such as an informed citizenry, opinion diversity, and tolerance for differing viewpoints.

9

CONCLUSION

With this dissertation, we advanced towards enabling responsible opinion formation in web search on debated topics. To do so, we have directed our efforts towards shedding light on the complex searcher-system interplay, as well as investigating approaches to mitigate the challenges arising from web search on debated topics. For that we adopted a searcher-centered perspective, seeking to understand variations in searchers' interaction choices across varying conditions. Specifically, we investigated how search interactions were shaped by distinct exposure bias (i.e., over-representation of one viewpoint over others on the SERP) and viewpoint transparency conditions, varying user characteristics, and nudging and boosting interventions aimed at supporting unbiased and diligent search behavior.

In this concluding chapter, we revisit our research questions to summarize and contextualize the findings and insights. We describe the technical and methodological implications, as well as the limitations of this dissertation. Along the way, we indicate directions for future research. Lastly, we discuss the ethical implications and broader challenges inherent to this research.

9.1. SUMMARY OF FINDINGS

With the findings and discussions presented in the four parts of this dissertation, we advanced the understanding of the searcher-system interplay (Part I), investigated the risks and benefits of interventions to guide search behavior (Part II) and empower searchers (Part III) to engage in unbiased and thorough search behavior, and outlined research challenges, providing a research agenda to advance the journey towards enabling responsible opinion formation in web search on debated topics (Part IV).

PART I: UNDERSTANDING THE SEARCHER-SYSTEM INTERPLAY

In the first part of this dissertation, we aimed to gain a comprehensive understanding of the dynamics of search on debated topics. Specifically, we sought to understand *how characteristics of the searcher and search system shape search on debated topics*

($\mathbf{RQ_{I}}$). For this, we conducted two user studies, exploring the relations between (i) factors inherent to the searcher and search system, (ii) search interaction, and (iii) post-search epistemic states.

The findings of the first user study, presented in Chapter 2, indicated that attributes of both the searcher and the search system shape search interactions. We observed that even though exposure bias considerably influenced search behavior, user interaction choices somewhat diminished the impact of attitude-opposing, but not attitude-supporting exposure bias. This reflects behavioral patterns of confirmation bias and corroborates other recent observations on the dominant role of searchers' interaction choices in the context of search on debated and political topics [302, 323].

We discovered that confirmation bias was most pronounced in searchers with strong prior attitudes, low prior topic knowledge, and those exposed to a SERP with a viewpoint bias aligned with their attitude. Exploring searchers' attitude change, we observed that search system exposure did not directly have an impact (as opposed to the findings on the search engine manipulation effect reported in [82]). Instead, participants' confirmation bias and their initial attitude strength affected attitude change. Higher confirmation bias was linked to reduced attitude change and individuals with strong prior attitudes were highly unlikely to change their attitude, regardless of other factors.

The findings of the second exploratory user study, presented in Chapter 3, indicate that increasing the ranking transparency by employing stance labels that indicate the stance of a search result (i.e., neutral, in favor, against) leads to a higher diversity of viewpoints searchers engage with and alleviates the impact of ranking bias, thereby promoting a more appropriate reliance on the search system and ranking.

Overall, it became apparent that to understand web search on debated topics and mitigate associated challenges, it is crucial to consider differences in searchers' pre-search epistemic states, particularly their attitude strength. Moreover, our findings in both chapters emphasize that searchers tend to fall back on means of simplifying the cognitively demanding endeavour of web search on debated topics, evident in their reliance on the opaque search ranking and exhibition of confirmation bias. The observed effect of stance labels leading to engagement with more diverse viewpoints, however, suggests that searchers want to explore diverse viewpoints and will do so if provided with measures that facilitate the process [117, 226, 320, 325].

Part II: Guiding Search Behavior

In this second part of the dissertation, we aimed to identify nudging interventions that mitigate searchers' confirmation bias. Particularly, we wanted to understand *whether we can guide individuals with warning labels and obfuscations to engage in unbiased search behavior on debated topics without harming their autonomy* ($\mathbf{RQ_{II}}$). To address this question, we conducted two user studies, investigating the effect of warning labels and obfuscations on searchers' engagement with attitude-confirming search results.

In the first study, presented in Chapter 4, we found that warning labels with obfuscations reduce engagement with targeted search results. However, the intervention reduced engagement even when applied to search results that were not attitude-confirming but randomly selected. This suggests that warning labels with obfuscations are a powerful tool to steer user behavior which could jeopardize user autonomy, raising concerns that it manipulates rather than guides user behavior.

We conducted a follow-up user study to gain further insights into the benefits and risks of warning labels and obfuscations to mitigate searchers' confirmation bias. With this study, presented in Chapter 5, we investigated the effect of warning labels and obfuscations separately. We discovered that obfuscations hinder searchers from detecting incorrect applications of the intervention, confirming our concerns that they harm user autonomy. Warning labels without obfuscations, however, motivated searchers to actively choose to decrease engagement with attitude-confirming search results while not preventing them from detecting incorrect applications to randomly selected search results.

The findings of both studies highlight the risk of obfuscations to guide search behavior in harming user autonomy, emphasizing that obfuscation and potentially other automatic nudging elements should be avoided in this context. Instead, interventions that aim at strengthening human cognitive skills and agency in engaging with information on debated topics would be preferable.

PART III: EMPOWERING THE SEARCHER

In an attempt to overcome some of the shortcomings of nudging interventions, identified in the previous part, in Part III we shift the focus towards autonomy-preserving boosting interventions that aim at empowering users to overcome challenges in web search on debated topics. In particular, we aimed to understand *whether we can empower individuals to engage in unbiased, as well as diligent search behavior, with interventions that boost their intellectual humility* (**RQ**$_{\mathbf{III}}$).

We approached this research question in two steps. First, in Chapter 6, we presented a theoretical account of the role of cognitive biases throughout the information search process, and the potential of IH-boosts in this context.

In the second step, presented in Chapter 7, we conducted a user study to test whether this potential would translate into unbiased and diligent search behavior in practice. While we identified three interventions that boost self-reported IH, we did not find an effect of the interventions on search behavior. Yet, explorations revealed that whether searchers approach the task with IH-related intentions affects their search behavior. Moreover, we observed that searchers' with higher levels of IH have more rewarding search experiences.

From both chapters, we derive that IH has great potential in supporting unbiased and diligent engagement with debated topics. However, our findings indicate that for the IH-boosting interventions to be effective, they likely require a more direct integration in the search process and should be combined with approaches that provide transparency of the viewpoint ranking.

Part IV: Towards Supporting Responsible Opinion Formation

In the final part of this dissertation, we adopted a broader perspective, aiming to inspire and facilitate future research efforts in striving to enable responsible opinion formation in web search on debated topics. To this end, we sought to understand *what the challenges and research opportunities towards supporting web search on debated topics and promoting responsible opinion formation are* (**RQ$_{\text{IV}}$**).

To answer this question, we built on perspectives from digital humanism and conducted an extensive interdisciplinary literature review. We provide a discussion of foundational challenges and research opportunities, for instance regarding value-tensions that emerge in this context. Further, we identified a range of challenges inherent to the searcher as the primary yet multifaceted stakeholder in web search on debated topics, such as understanding the impact of diverse characteristics, attitudes, knowledge levels, competencies, and cognitive biases. We outline additional research opportunities pertaining to the search engine as the non-neutral system that determines what information searchers are exposed to. For instance, research is needed to understand how to meaningfully re-rank search results for increased viewpoint diversity. Moreover, we identify research challenges arising from the entangled interplay between searcher and search engine, such as understanding how to leverage the search interface to increase ranking transparency and encourage thorough search behavior. Lastly, we point towards broader challenges that need addressing in the context of web search on debated topics, for instance, conducting insightful user studies, considering alternative search paradigms, and dealing with misinformation.

Through this undertaking, we provide a foundation for future research efforts toward a more enriching, unbiased, and trustworthy experience when searching for information on debated topics.

9.2. Implications, Reflections, and Methodological Insights

The findings from research projects in this dissertation have implications for the pursuit of enabling responsible opinion formation in web search, as well as various research communities, which we will discuss in the following. To benefit future research efforts, we will also share the reflections and methodological insights we gained throughout the process.

9

9.2.1. Towards Enabling Responsible Opinion Formation

Our combined observations highlight that search on debated topics, characterized by diverse perspectives, the heightened risk of interacting searcher and search engine biases, and its consequential nature, impacting individuals and society at large, warrants dedicated research effort.

Throughout our research journey, we learned that to tackle challenges associated with web search on debated topics approaches that address issues in isolation (e.g., mitigating searchers' confirmation bias) are not sufficient. Instead, approaches should aim to encourage unbiased and diligent search interactions, increasing searchers' informedness, interest in, and tolerance for diverse viewpoints. Regarding these objectives, the current trend towards search interfaces that provide answers ever more seamlessly to the searcher, e.g., by providing direct responses with featured snippets, appears to be detrimental. From our observations of how interventions aimed at guiding search behavior run the risk of manipulating searchers, we posit that interventions should follow the strict principle of increasing transparency to and choice for the user, augmenting and complementing human cognitive skills instead of manipulating or herding them [85, 398].

To achieve increased transparency and choice, future research should investigate approaches to provide additional information about the search results (e.g., epistemic cues [210, 325]), yet in a manner that does not increase information overload. This direction for future work is supported by our findings on how reflective nudges consisting of warning labels, and stance labels that provide transparency of the viewpoint ranking effectively increase the diversity of viewpoints searchers engage with without harming their autonomy. While we did not find evidence for an effect of intellectual humility boosts on search behavior, exploratory insights suggest that future research should investigate them further, for instance, integrating them more directly into the search process (e.g., as pop-up messages) and combining them with stance labels to provide transparency of the viewpoint ranking.

Reflecting on the applied methods of our user studies, we primarily conducted strictly controlled, confirmatory user studies to test preregistered hypotheses, aiming to gain valid insights. Yet, we often noted high levels of noise in our dependent variables, indicating that other factors than the ones we considered and manipulated played a role in shaping participants' interaction behavior. This general observation underscores that for research focusing on socio-technical issues, it is difficult to isolate single factors while ignoring others. This calls for more exploratory approaches to advance a comprehensive understanding of human information interaction, in addition to controlled confirmatory studies [315].

9.2.2. Behavioral Change Interventions

We conducted user studies testing the effect of reflective and automatic nudging interventions (i.e., warning labels and obfuscations), and boosting interventions (i.e., intellectual humility boosts) on search behavior.

We found that obfuscations prevent users from detecting incorrect applications of the interventions, restricting their agency and harming their autonomy. Consequently, we infer that automatic nudging interventions run the risk of manipulating rather than guiding user behavior. This could be the case even if interventions were designed to be transparent, provided that users do not sufficiently engage with the transparent element, but change their information behavior in response to the automatic element. Overall, our results suggest caution in applying nudging interventions that directly steer user behavior, particularly if the *beneficial behavior* towards which users will be nudged cannot be clearly defined due to complex context and user dependencies (e.g., prior knowledge, attitude strength). However, we want to emphasize that no user interface design is neutral. Instead, every design is grounded in choice architectures that direct user behavior, which on the web tend to be guided by commercial interests [178, 180].

Boosting interventions promise to overcome the shortcomings of nudging interventions, preserving user autonomy by improving searchers' competencies rather than directly steering their behavior. However, outcomes emerging from our work point towards the challenge of modifying behavior with such autonomy-preserving interventions in highly familiar online environments that individuals learned to rely on through past interactions. Thus, we infer that boosting interventions would need to be integrated into the search process to overcome behavioral patterns. Additionally, they should be combined with other measures that support appropriate reliance on the system functions (e.g., stance labels). Recent research by Bink and Elsweiler [32], who investigated the effect of information literacy boosting interventions that were more directly integrated into the search process, shows that such interventions effectively encourage searchers with neutral attitudes to invest more effort in searching on debated topics. Whether these findings would extend to both intellectual humility boosts and opinionated searchers has yet to be investigated.

In general, our findings illustrate that the benefits and risks of interventions need to be carefully investigated in the target environment, monitoring immediate and long-term effects. This resonates with the Freiling et al.'s [91] perspective who stress the need to consider the intricate nature of the issues that behavioral interventions intend to tackle, as well as the broader (ethical) implications inherent to envisioned interventions. For instance, the authors discuss how interventions aimed at guiding online information behavior during the COVID-19 pandemic may have caused effects beyond the information interaction, such as diminishing trust in science and governmental institutions.

9.2.3. Human Information Interaction and Retrieval

Throughout our studies, we did not find evidence for a direct effect of exposure bias on searchers' attitudes, not aligning with the observations of the *search engine manipulation effect* reported in [82]. Reflecting on this finding, we argue that attitudes on debated topics, particularly strong attitudes, tend to be rooted in stable values [140] and might thus be less directly affected by exposure bias than the candidate-related voting decisions investigated by Epstein et al. [82].

Moreover, we observed that confirmation bias drives user interaction to somewhat diminish the impact of attitude-opposing exposure bias. This corroborates recent observations on the importance of users' interaction choices in web search on political and debated topics [302, 323]. While these observations could suggest that exposure bias might not pose a risk to responsible opinion formation on debated topics, our findings point towards an alternate risk, namely attitude-supporting exposure bias which amplifies searchers' interaction biases. Relevance criteria that aim at increasing the satisfaction of the individual user during search on debated topics would likely result in attitude-confirming exposure bias, which would hinder engagement with diverse viewpoints and responsible opinion formation.

The risk of reinforced interaction biases could be mitigated by increasing the ranking transparency. In this regard, our findings on how displaying stance labels with search results increases the diversity of viewpoints searchers engage with and counters the effect of viewpoint-biased exposure are promising. These findings underscore that searchers indeed want to explore diverse viewpoints when provided with the cues to differentiate between search results, but have been conditioned to rely on opaque search engine rankings [81, 117, 226, 320, 325].

9.2.4. USER MODELLING AND ADAPTION

The findings of our user studies consistently emphasize the importance of diverse user characteristics, particularly their pre-search epistemic states, in shaping both search interaction on debated topics, as well as the effect of interventions to support unbiased and thorough search behavior. For interventions to effectively support productive searches on debated topics, they would thus likely need to be personalized to benefit all searchers. The important role of user characteristics that we observed emphasizes that understanding the effects of socio-technical systems and mitigating associated risks requires user-centered research and design.

9.3. LIMITATIONS AND DIRECTIONS FOR FUTURE WORK

The research conducted in the scope of this dissertation focused on the searcher and their interplay with search results during web search on debated topics. Thus, we mostly disregarded the challenges of web search on debated topics associated with the search system (e.g., viewpoint detection, ranking algorithms for viewpoint diversity). Instead, this research relied heavily on controlled and crowdsourced user studies to advance the understanding of searchers' interactions with search results under various conditions.

Reflecting on our research approach, we acknowledge several limitations. For all user studies, we created artificial scenarios that prompted participants whom we recruited via a crowd-working platform to search for information on a debated topic, disregarding whether they were interested in learning about the assigned topic. We presented them with mock search interfaces that mimicked conventional search systems, yet all search results were derived from a set of preselected, viewpoint-annotated search results. To control for or investigate ranking effects, we applied ranking templates of selected viewpoint orders, rather than considering relevance to participants' queries. All of these circumstances might have caused the participants to interact with the search page in a manner different from their usual search behavior when searching for information on a debated topic out of a genuine information need. Despite the potential shortcomings, we opted for controlled studies that rely on crowdworkers, as this allowed for big sample sizes to obtain confirmatory insights into how individual factors affected search behavior.

To apply viewpoint ranking templates and measure interactions with different viewpoints, we required viewpoint-annotated search results. For that, we relied on ternary stance labels (i.e., neutral, in favor, against), which do not sufficiently reflect the complexity of viewpoints and viewpoint differences [68]. Moreover, we investigated singular search sessions on one debated topic in which we focused primarily on searchers' clicks on search results, disregarding other parts of the information search process, such as querying in our analyses. We did not investigate how the search interactions affected searchers' beyond attitude change and knowledge gain, e.g., whether they affected practical decision-making. More comprehensive studies are needed to gain insights into longer-term and topic-independent behavioral patterns and effects of the tested interventions, considering the full information search process and applying viewpoint taxonomies that better capture the complexity of diverse viewpoints. In light of recent insights into the changing role of active search compared to passive exposure on social media platforms (see e.g., [125, 233]), a need for qualitative user studies to gain a complete understanding of opinion formation on the web becomes evident.

Reflecting on our findings, we want to provide some pointers for direct future work to follow up on our research. To advance the understanding of searchers' characteristics in web search on debated topics, future research should investigate how nuanced variances in factors related to their attitudes, such as *attitude importance*, *moral conviction*, or *attitude certainty*, factor into their interactions with search results and reaction to behavioral interventions.

In our studies, we identified that an important factor in shaping search interactions appears to be searchers' reliance on the opaque search ranking. This reliance may be attributed to the design of search interfaces that aim to provide the right information to satisfy an information need with minimal effort required from the searcher [320]. Further research efforts should be directed towards identifying means of supporting more appropriate reliance, for instance, by introducing more transparency to the interface design to reveal the limits of search engine processes when dealing with queries on debated topics (i.e., viewpoint ranking bias) [147].

To support searchers comprehensively in overcoming their own biases and increasing their tolerance for diverse viewpoints when engaging with debated topics, IH-boosting interventions were the most promising of the interventions we investigated. However, we did not find evidence for an effect on search interactions, potentially due to administering them prior to the search task and not combining them with interventions that increase the transparency of the viewpoint ranking to achieve appropriate reliance. To better understand the potential of such interventions, future research should investigate IH-boosting interventions further, for instance when more directly integrated into the search process (see for example [32]) and combined with measures that enhance search engine transparency, such as search result stance labels.

9.4. Ethical Considerations and Challenges

As this dissertation investigates issues concerning access to information, individual opinions, and societal well-being, it is crucial to identify and consider potential ethical pitfalls and challenges.

A foundational challenge consists of the value tensions that emerge when attempting to design search systems and interfaces that better support individual and societal needs and values [44, 48, 330]. For search on debated topics, for instance, relevance to user needs might not be aligned with relevance to societal and democratic values, necessitating an in-depth discussion of value trade-offs. Such discussion should be grounded in pertinent human rights such as the *freedom of opinion and expression* [161, 361].

Further, a prevalent risk of developing interventions to mitigate risks in human information behavior is the paternalistic inclination to perceive users as patients in need of fixing and protection from their reasoning errors and inertia, rather than individuals capable of learning to navigate potential risks autonomously [91, 100]. Affirming this criticism, we have identified several shortcomings and risks of automatic nudging interventions that directly steer user behavior, including the risk of misuse for malicious purposes. Conversely, reflective nudges effectively prompted searchers to choose to engage with diverse viewpoints, rather than exhaust their processing capabilities. Reflecting on these insights, we advocate against the application of automatic nudges that directly steer behavior to mitigate searchers' confirmation bias. However, our findings also revealed how environments of information overload, coupled with the seamless and opaque design of search interfaces provoke human bias and inertia, posing a considerable challenge to unbiased and diligent search on debated topics, irrespective of whether searchers intend to gain a well-rounded knowledge on the topic. Thus, even in the absence of interventions, search engines steer information behavior and provoke certain behavioral patterns, acting as algorithmic curators that determine which information individuals can access [342, 361, 372]. Consequently, there is a need for approaches to support unbiased and diligent search behavior that guarantee or even enhance user agency by providing and facilitating choice and increasing transparency to the user rather than attempting to directly steer their behavior [147, 320, 398]. Yet, potential approaches require thorough investigation to avoid unexpected consequences and tensions that are prevalent in entangled socio-technical systems [91, 248].

Our findings suggest potential benefits of personalized interventions to support unbiased and diligent search behaviour in diverse searchers, for instance, adapting to their pre-search epistemic states. Yet, this would require the elicitation and storage of sensitive personal information, such as individuals' opinions on socio-scientific topics and political beliefs, raising concerns about the practical implementation and user privacy and GDPR compliance. To mitigate these risks, personalization should be approached in a privacy-aware manner, guaranteeing that users comprehend and have control over the factors influencing their user model, as well as its impact on the information environment [344]. Alternatively, interventions that do not require personalization (e.g., boosting interventions, stance labels) should be prioritized.

Lastly, we want to emphasize that the challenges discussed in this section should be considered in perspective to the status quo: While having become a vital infrastructure for information access, most search engines are not run as public goods but instead by private industry, likely prioritizing profit motives over societal well-being when deciding over an individual's information access [105, 199, 246, 320, 330, 361]. Considering the risks of this status quo, sticking to it is not a viable option. Despite the tensions and challenges that lie ahead, we should thus continue this line of research, striving for responsible opinion formation in web search.

9.5. CONCLUDING REMARKS

Recent years have seen heated debates on how to restructure societies and adjust the ways we live in response to a changing world. For a healthy democratic culture, individuals would ideally engage with different perspectives in such debates to form opinions responsibly. Yet, this can be cognitively and emotionally challenging. Moreover, the rapidly shifting information landscape due to the uptake of digital technology might hinder rather than support engagement with diverse viewpoints—a risk that necessitates investigation to gain a comprehensive understanding and identify effective mitigation approaches.

This dissertation centers on interaction with debated topics via web search, which has become one of the primary gateways to information. By disentangling the searcher-system interplay and investigating approaches to support searchers in overcoming the challenges arising from web search on debated topics, we have advanced the pursuit of enabling responsible opinion formation in web search. Our findings reveal challenges, such as disrupting unproductive behavioral patterns without harming user autonomy, and opportunities, such as increasing viewpoint-ranking transparency for interactions with more diverse viewpoints.

Overall, our findings emphasize that the greatest challenges arise from the entangled user-system interplay, reflecting a core characteristic of many socio-technical issues. This is illustrated by phenomena such as learned reliance on the search ranking that hinders the detection of viewpoint-biased rankings. Moreover, supporting responsible opinion formation in web search does not only require understanding and mitigating risks. It also requires a discussion of values in tension, such as profit-driven information environments and individual user satisfaction vs. societal well-being, support in navigating information overload vs. user autonomy, and personalization vs. neutrality and privacy, to name only a few. Thus, we find ourselves at a pivotal crossroads, requiring decisions that will determine the future of information access. Meanwhile, search engine functions change rapidly and non-transparently, as private technology companies make such decisions, bypassing public deliberation and democratic legitimization.

Throughout this dissertation journey, we continuously uncovered additional levels of complexity inherent to web search on debated topics. Yet, it is essential to avoid easy fixes and, instead, acknowledge and navigate through these levels of complexity, given that we address issues concerning information access and societal well-being. Consequently, the pursuit of enabling responsible opinion formation in web search is only beginning, and we hope that the work conducted in the scope of this dissertation motivates others to join. To make progress, we need interdisciplinary efforts involving researchers with diverse backgrounds, such as computer scientists, information and communication scientists, psychologists, philosophers, and lawyers. We believe that through such efforts, web search platforms can become spaces that encourage and empower individuals to acquire well-rounded knowledge on debated topics and foster a sense of collectivity rather than division in tackling societal challenges.

Glossary

Below, we provide an alphabetical overview of key terms, along with their explanations and definitions, as used in this dissertation.

Boosting. Interventions that aim to cultivate enduring user competencies, retaining their effect even after being presented to the individual [132, 210].

Cognitive biases. Systematic deviations in judgment and decision-making from what would be expected based on rational decision-making models [351].

Confirmation bias. A cognitive bias that occurs when individuals favor information that confirms preexisting beliefs while dismissing or discounting information that opposes those beliefs [244].

Cognitive style. An individual's tendency to rely more on analytic, effortful or intuitive, effortless thinking, closely related to their *need for cognition* [46, 90].

Debated topics. Topics that are characterized by an ongoing discussion among groups of individuals with different perspectives and opinions, often tied to conflicting values or interests, and lacking a straightforward resolution [308].

Digital Humanism. An initiative that advocates for analyzing and reflecting on human-technology relationships, promoting human-centered technology that prioritizes improved lives and societal progress over economic growth [371].

Elaboration likelihood model. A theoretical framework that distinguishes between the *peripheral* and the *central* route of processing persuasive interventions [262]. The peripheral route relies on simple, non-argumentative cues to evoke intuitive and unconscious reactions, while the central route engages the critical thinking skills of the recipient to evaluate the presented arguments.

Epistemic states. Individual states (i.e., temporary conditions) that are related to knowledge and opinions, such as attitudes on a topic or topic knowledge [299].

Exploratory search. Search that is focused on investigation and learning, rather than retrieving facts to answer a particular question (*lookup search*) [215].

Exposure bias. Occurs when search engines disproportionately highlight one viewpoint over others on their result pages [299, 302].

Featured snippet. A direct answer to a search query displayed at the top of the search engine results page, extracted from a webpage, and linked back to that page.

Functional fixedness. A phenomenon wherein individuals experience constraints to use a tool in unfamiliar ways [4, 79].

Intellectual humility. A relatively stable individual trait, as well as context-dependent state, encompassing the competencies to recognize the limits of one's knowledge and be aware of the fallibility of one's opinions and beliefs [267].

Need for cognition. An individual's tendency to engage in effortful, analytical thinking to organize their experience meaningfully [62].

Nudging. Interventions to subtly guide individuals to make decisions that are considered to be beneficial for them, without restricting possible choices, e.g., by setting defaults, creating friction, or suggesting alternatives [49, 337].

Position bias. Searchers' tendency to engage primarily with items that appear in the higher ranks of a search engine result page [18].

Receptiveness to opposing views. The extent to which individuals are willing to access, consider, and evaluate opposing views in an impartial manner [229].

Responsible opinion formation. Occurs when individuals actively seek out and diligently engage with diverse viewpoints on a topic, critically evaluating them to gain well-rounded knowledge before forming or changing their opinion [177, 260].

Seamful design. A concept in human-computer interaction design that emphasizes configurability and transparency, revealing complexity, ambiguity, or inconsistency, in contrast to *seamless design*, which prioritizes clarity, simplicity, ease of use, and consistency [147].

Search engine manipulation effect. A phenomenon that occurs when searchers adopt the viewpoint that is over-represented on a search engine result page [82].

Search engine results page (SERP). The page displayed by a search engine in response to a query. It typically contains a list of search results and often additional elements such as featured snippets and ads.

Stance labels. Labels that show the stance of a search result on a specific debated topic (e.g., neutral, in favor, against).

BIBLIOGRAPHY

[1] Eugene Agichtein, Eric Brill, and Susan Dumais. "Improving Web Search Ranking by Incorporating User Behavior Information". In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '06. New York, NY, USA: Association for Computing Machinery, 2006, pp. 19–26. ISBN: 978-1-59593-369-0. DOI: 10.1145/1148170.1148177.

[2] Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Ieong. "Diversifying search results". en. In: *Proceedings of the Second ACM International Conference on Web Search and Data Mining - WSDM '09*. Barcelona, Spain: ACM Press, 2009, p. 5. ISBN: 978-1-60558-390-7. DOI: 10.1145/1498759.1498766. URL: http://portal.acm.org/citation.cfm?doid=1498759.1498766 (visited on 07/13/2021).

[3] Yamen Ajjour, Pavel Braslavski, Alexander Bondarenko, and Benno Stein. "Identifying Argumentative Questions in Web Search Logs". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '22. Madrid, Spain: Association for Computing Machinery, 2022, pp. 2393–2399. ISBN: 9781450387323. DOI: 10.1145/3477495.3531864.

[4] Marwah Alaofi, Luke Gallagher, Dana Mckay, Lauren L. Saling, Mark Sanderson, Falk Scholer, Damiano Spina, and Ryen W. White. "Where Do Queries Come From?" In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '22. Madrid, Spain: Association for Computing Machinery, 2022, pp. 2850–2862. ISBN: 9781450387323. DOI: 10.1145/3477495.3531711.

[5] Abeer ALDayel and Walid Magdy. "Stance Detection on Social Media: State of the Art and Trends". en. In: *Information Processing & Management* 58 (July 2021), p. 102597. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2021.102597.

[6] Mark Alfano, Kathryn Iurino, Paul Stey, Brian Robinson, Markus Christen, Feng Yu, and Daniel Lapsley. "Development and Validation of a Multi-Dimensional Measure of Intellectual Humility". In: *PLOS ONE* 12 (2017), e0182950. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0182950.

[7] Ahmed Allam, Peter Johannes Schulz, and Kent Nakamoto. "The Impact of Search Engine Selection and Sorting Criteria on Vaccination Beliefs and Attitudes: Two Experiments Manipulating Google Output". In: *Journal of Medical Internet Research* 16 (2014), e2642. DOI: 10.2196/jmir.2642.

[8] Jisun An, Daniele Quercia, Meeyoung Cha, Krishna Gummadi, and Jon Crowcroft. "Sharing Political News: The Balancing Act of Intimacy and Socialization in Selective Exposure". In: *EPJ Data Science* 3 (2014), p. 12. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-014-0012-2.

[9] Evangelia Anagnostopoulou, Babis Magoutas, Efthimios Bothos, Johann Schrammel, Rita Orji, and Gregoris Mentzas. "Exploring the Links Between Persuasion, Personality and Mobility Types in Personalized Mobility Applications". en. In: *Persuasive Technology: Development and Implementation of Personalized Technologies to Change Attitudes and Behaviors*. Ed. by Peter W. de Vries, Harri Oinas-Kukkonen, Liseth Siemons, Nienke Beerlage-de Jong, and Lisette van Gemert-Pijnen. Vol. 10171. Cham: Springer International Publishing, 2017, pp. 107–118. DOI: 10.1007/978-3-319-55134-0_9.

[10] Alessia Antelmi, Delfina Malandrino, and Vittorio Scarano. "Characterizing the Behavioral Evolution of Twitter Users and The Truth Behind the 90-9-1 Rule". In: *Companion Proceedings of The 2019 World Wide Web Conference*. WWW '19. San Francisco, USA: Association for Computing Machinery, 2019, pp. 1035–1038. ISBN: 9781450366755. DOI: 10.1145/3308560.3316705.

[11] Theo Araujo, Jef Ausloos, Wouter van Atteveldt, Felicia Loecherbach, Judith Moeller, Jakob Ohme, Damian Trilling, Bob van de Velde, Claes De Vreese, and Kasper Welbers. “OSD2F: An open-source data donation framework”. In: *Computational Communication Research* 4.2 (2022), pp. 372–387.

[12] Jennifer J. Argo and Kelley J. Main. “Meta-Analyses of the Effectiveness of Warning Labels”. en. In: *Journal of Public Policy & Marketing* 23 (Sept. 2004), pp. 193–208. ISSN: 0748-6766, 1547-7207. DOI: 10.1509/jppm.23.2.193.51400.

[13] Charles Arthur. *What is the 1% rule?* July 2006. URL: https://www.theguardian.com/technology/2006/jul/20/guardianweeklytechnologysection2.

[14] Abolfazl Asudeh, H. V. Jagadish, Julia Stoyanovich, and Gautam Das. “Designing Fair Ranking Schemes”. en. In: *Proceedings of the 2019 International Conference on Management of Data*. Amsterdam Netherlands: ACM, June 2019, pp. 1259–1276. ISBN: 978-1-4503-5643-5. DOI: 10.1145/3299869.3300079. (Visited on 07/13/2021).

[15] Kumaripaba Athukorala, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. “Is Exploratory Search Different? A Comparison of Information Search Behavior for Exploratory and Lookup Tasks”. In: *Journal of the Association for Information Science and Technology* 67 (2016), pp. 2635–2651. ISSN: 2330-1643. DOI: 10.1002/asi.23617.

[16] Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. “Stance Detection with Bidirectional Conditional Encoding”. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. New York, NY, USA: Association for Computing Machinery, 2016, pp. 876–885.

[17] Sandeep Avula, Gordon Chadwick, Jaime Arguello, and Robert Capra. “Searchbots: User engagement with chatbots during collaborative search”. In: *Proceedings of the 2018 conference on human information interaction & retrieval*. 2018, pp. 52–61.

[18] Leif Azzopardi. “Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval”. In: *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. Canberra ACT Australia: ACM, 2021, pp. 27–37. ISBN: 978-1-4503-8055-3. DOI: 10.1145/3406522.3446023.

[19] Christian Baden, Neta Kligler-Vilenchik, and Moran Yarchi. “Hybrid Content Analysis: Toward a Strategy for the Theory-driven, Computer-assisted Classification of Large Text Corpora”. en. In: *Communication Methods and Measures* 14.3 (July 2020), pp. 165–183. ISSN: 1931-2458, 1931-2466. DOI: 10.1080/19312458.2020.1803247. (Visited on 07/13/2021).

[20] Christian Baden and Nina Springer. “Com(ple)menting the news on the financial crisis: The contribution of news users’ commentary to the diversity of viewpoints in the public debate”. en. In: *European Journal of Communication* 29.5 (Oct. 2014), pp. 529–548. ISSN: 0267-3231, 1460-3705. DOI: 10.1177/0267323114538724. (Visited on 07/13/2021).

[21] Christian Baden and Nina Springer. “Conceptualizing viewpoint diversity in news discourse”. en. In: *Journalism* 18.2 (Feb. 2017), pp. 176–194. ISSN: 1464-8849, 1741-3001. DOI: 10.1177/1464884915605028. (Visited on 07/13/2021).

[22] Ricardo Baeza-Yates. “Bias on the Web”. In: *Communications of the ACM* 61 (2018), pp. 54–61.

[23] Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. “Exposure to opposing views on social media can increase political polarization”. In: *Proceedings of the National Academy of Sciences* 115.37 (2018), pp. 9216–9221.

[24] Judit Bar-Ilan, Kevin Keenoy, Mark Levene, and Eti Yaari. “Presentation bias is significant in determining user preference for search results—A user study”. In: *Journal of the American Society for Information Science and Technology* 60.1 (2009), pp. 135–149.

[25] Francesco Barile, Tim Draws, Oana Inel, **Alisa Rieger**, Shabnam Najafian, Amir Ebrahimi Fard, Rishav Hada, and Nava Tintarev. “Evaluating Explainable Social Choice-Based Aggregation Strategies for Group Recommendation”. In: *User Modeling and User-Adapted Interaction* 34 (2024), pp. 1–58. ISSN: 1573-1391. DOI: 10.1007/s11257-023-09363-0.

[26] Francesco Barile, Shabnam Najafian, Tim Draws, Oana Inel, **Alisa Rieger**, Rishav Hada, and Nava Tintarev. "Toward Benchmarking Group Explanations: Evaluating the Effect of Aggregation Strategies versus Explanation." In: *Perspectives@ RecSys*. 2021.

[27] Michael A Beam. "Automating the news: How personalized news recommender system design choices impact news reception". In: *Communication Research* 41.8 (2014), pp. 1019–1041.

[28] Nicholas J. Belkin. "Anomalous states of knowledge as a basis for information retrieval". In: *Canadian Journal of Information Science* 5.1 (1980), pp. 133–143.

[29] Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. "Detecting Personality Traits Using Eye-Tracking Data". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–12. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300451.

[30] Abraham Bernstein, Claes H. de Vreese, Natali Helberger, Wolfgang Schulz, and Katharina A. Zweig. "Diversity, Fairness, and Data-Driven Personalization in (News) Recommender System (Dagstuhl Perspectives Workshop 19482)". In: *Dagstuhl Manifestos* 9 (2020), pp. 117–124. ISSN: 2193-2433. DOI: 10.4230/DagRep.9.11.117.

[31] Michael S Bernstein, Jaime Teevan, Susan Dumais, Daniel Liebling, and Eric Horvitz. "Direct answers for search queries in the long tail". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2012, pp. 237–246.

[32] Markus Bink and David Elsweiler. "Balancing Act: Boosting Strategies for Informed Search on Controversial Topics". In: *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*. 2024, pre–print.

[33] Markus Bink, Sebastian Schwarz, Tim Draws, and David Elsweiler. "Investigating the Influence of Featured Snippets on User Attitudes". en. In: *ACM SIGIR Conference on Human Information Interaction and Retrieval*. CHIIR '23. New York, NY, USA: ACM, 2023. DOI: 10.1145/3576840.3578323.

[34] Markus Bink, Steven Zimmerman, and David Elsweiler. "Featured Snippets and Their Influence on Users' Credibility Judgements". In: *ACM SIGIR Conference on Human Information Interaction and Retrieval*. CHIIR '22. Regensburg, Germany: Association for Computing Machinery, 2022, pp. 113–122. ISBN: 9781450391863. DOI: 10.1145/3498366.3505766.

[35] M. José Blanca Mena, Rafael Alarcón Postigo, Jaume Arnau Gras, Roser Bono Cabré, and Rebecca Bendayan. "Non-Normal Data: Is ANOVA Still a Valid Option?" In: (2017). ISSN: 0214-9915.

[36] Sina Blassnig, Eliza Mitova, Nico Pfiffner, and Michael V Reiss. "Googling referendum campaigns: analyzing online search patterns regarding Swiss direct-democratic votes". In: *Media and Communication* 11.1 (2023), pp. 19–30.

[37] Luc Boltanski and Laurent Thévenot. *On justification: Economies of worth*. Vol. 27. Princeton University Press, 2006.

[38] Alexander Bondarenko, Maik Fröbe, Johannes Kiesel, Shahbaz Syed, Timon Gurcke, Meriem Beloucif, Alexander Panchenko, Chris Biemann, Benno Stein, Henning Wachsmuth, Martin Potthast, and Matthias Hagen. "Overview of Touché 2022: Argument Retrieval". In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Ed. by Alberto Barrón-Cedeño, Giovanni Da San Martino, Mirko Degli Esposti, Fabrizio Sebastiani, Craig Macdonald, Gabriella Pasi, Allan Hanbury, Martin Potthast, Guglielmo Faggioli, and Nicola Ferro. Cham: Springer International Publishing, 2022, pp. 311–336. ISBN: 978-3-031-13643-6.

[39] Horatiu Bota, Ke Zhou, and Joemon M. Jose. "Playing Your Cards Right: The Effect of Entity Cards on Search Behaviour and Workload". In: *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*. CHIIR '16. Carrboro, North Carolina, USA: Association for Computing Machinery, 2016, pp. 131–140. ISBN: 9781450337519. DOI: 10.1145/2854946.2854967.

[40] Shauna M. Bowes, Thomas H. Costello, Caroline Lee, Stacey McElroy-Heltzel, Don E. Davis, and Scott O. Lilienfeld. "Stepping Outside the Echo Chamber: Is Intellectual Humility Associated With Less Political Myside Bias?" In: *Personality and Social Psychology Bulletin* 48 (2022), pp. 150–164. ISSN: 0146-1672. DOI: 10.1177/0146167221997619.

[41] Alejandra Bringas Colmenarejo, Luca Nannini, **Alisa Rieger**, Kristen M. Scott, Xuan Zhao, Gourab K Patro, Gjergji Kasneci, and Katharina Kinder-Kurlanda. "Fairness in Agreement With European Values: An Interdisciplinary Perspective on AI Regulation". In: *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. AIES '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 107–118. ISBN: 978-1-4503-9247-1. DOI: 10.1145/3514094.3534158.

[42] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making". In: *Proceedings of the ACM on Human-Computer Interaction* 5 (2021), 188:1–188:21. DOI: 10.1145/3449287.

[43] Katarzyna Budzynska and Chris Reed. "Advances in Argument Mining". en. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*. Florence, Italy: Association for Computational Linguistics, 2019, pp. 39–42. DOI: 10.18653/v1/P19-4008. URL: https://www.aclweb.org/anthology/P19-4008 (visited on 07/13/2021).

[44] Robin Burke. "Personalization, Fairness, and Post-Userism". In: *Perspectives on Digital Humanism* (2022), p. 145.

[45] Matthew K. Buttice and Walter J. Stone. "Candidates Matter: Policy and Quality Differences in Congressional Elections". In: *The Journal of Politics* 74 (2012), pp. 870–887. ISSN: 0022-3816. DOI: 10.1017/S0022381612000394.

[46] John T. Cacioppo, Richard E. Petty, and Chuan Feng Kao. "The Efficient Assessment of Need for Cognition". In: *Journal of Personality Assessment* 48 (1984), pp. 306–307. ISSN: 0022-3891, 1532-7752. DOI: 10.1207/s15327752jpa4803_13.

[47] John T. Cacioppo, Richard E. Petty, and Katherine J. Morris. "Effects of Need for Cognition on Message Evaluation, Recall, and Persuasion". In: *Journal of Personality and Social Psychology* 45.4 (1983), pp. 805–818. ISSN: 1939-1315(Electronic),0022-3514(Print). DOI: 10.1037/0022-3514.45.4.805.

[48] Cansu Canca. "Did you find it on the internet? ethical complexities of search engine rankings". In: *Perspectives on Digital Humanism* (2022), p. 135.

[49] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. "23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–15. ISBN: 978-1-4503-5970-2.

[50] Noel Carroll. "In Search We Trust: Exploring How Search Engines Are Shaping Society". In: *International Journal of Knowledge Society Research (IJKSR)* 5 (2014), pp. 12–27. DOI: 10.4018/ijksr.2014010102.

[51] Carlos Castillo and Brian D. Davison. "Adversarial Web Search". In: *Foundations and Trends® in Information Retrieval* 4 (2011), pp. 377–486. ISSN: 1554-0669, 1554-0677. DOI: 10.1561/1500000021.

[52] Jon Chamberlain, Udo Kruschwitz, and Orland Hoeber. "Scalable visualisation of sentiment and stance". In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. 2018.

[53] Laurianne Charrier, **Alisa Rieger**, Alexandre Galdeano, Amélie Cordier, Mathieu Lefort, and Salima Hassas. "The RoPE Scale: A Measure of How Empathic a Robot Is Perceived". In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2019, pp. 656–657. DOI: 10.1109/HRI.2019.8673082.

[54] Sihao Chen, Daniel Khashabi, Wenpeng Yin, Chris Callison-Burch, and Dan Roth. "Seeing Things from a Different Angle: Discovering Diverse Perspectives about Claims". en. In: (June 2019). arXiv: 1906.03538 [cs].

[55] Sijing Chen, Lu Xiao, and Akit Kumar. "Spread of Misinformation on Social Media: What Contributes to It and How to Combat It". In: *Computers in Human Behavior* 141 (2023), p. 107643. ISSN: 0747-5632. DOI: 10.1016/j.chb.2022.107643.

[56] Tong Chen, Hongzhi Yin, Guanhua Ye, Zi Huang, Yang Wang, and Meng Wang. "Try this instead: Personalized and interpretable substitute recommendation". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2020, pp. 891–900.

[57] Haibin Cheng and Erick Cantú-Paz. "Personalized Click Prediction in Sponsored Search". In: *Proceedings of the Third ACM International Conference on Web Search and Data Mining*. WSDM '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 351–360. ISBN: 978-1-60558-889-6. DOI: 10.1145/1718487.1718531.

[58] Markus Christen, Mark Alfano, and Brian Robinson. "A Cross-Cultural Assessment of the Semantic Dimensions of Intellectual Humility". In: *AI & SOCIETY* 34 (2019), pp. 785–801. ISSN: 1435-5655. DOI: 10.1007/s00146-017-0791-7.

[59] Massimiliano Ciaramita, Vanessa Murdock, and Vassilis Plachouras. "Online Learning from Click Data for Sponsored Search". In: *Proceedings of the 17th International Conference on World Wide Web*. WWW '08. New York, NY, USA: Association for Computing Machinery, 2008, pp. 227–236. ISBN: 978-1-60558-085-2. DOI: 10.1145/1367497.1367529.

[60] Jason K. Clark and Duane T. Wegener. "Chapter Four - Message Position, Information Processing, and Persuasion: The Discrepancy Motives Model". In: *Advances in Experimental Social Psychology*. Ed. by Patricia Devine and Ashby Plant. Vol. 47. Academic Press, 2013, pp. 189–232. DOI: 10.1016/B978-0-12-407236-7.00004-8.

[61] Katherine Clayton, Spencer Blair, Jonathan A. Busam, Samuel Forstner, John Glance, Guy Green, Anna Kawata, Akhila Kovvuri, Jonathan Martin, Evan Morgan, Morgan Sandhu, Rachel Sang, Rachel Scholz-Bright, Austin T. Welch, Andrew G. Wolff, Amanda Zhou, and Brendan Nyhan. "Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media". In: *Political Behavior* 42 (2020), pp. 1073–1095. ISSN: 1573-6687. DOI: 10.1007/s11109-019-09533-0.

[62] Arthur R. Cohen, Ezra Stotland, and Donald M. Wolfe. "An Experimental Investigation of Need for Cognition." In: *The Journal of Abnormal and Social Psychology* 51.2 (1955), pp. 291–294. ISSN: 0096-851X. DOI: 10.1037/h0042761. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0042761 (visited on 04/03/2021).

[63] Savia A. Coutinho. "The Relationship between the Need for Cognition, Metacognition, and Intellectual Task Performance". In: *Educational Research and Reviews* 1 (2006), pp. 162–164. ISSN: 1990-3839. DOI: 10.5897/ERR.9000373.

[64] Stephen Cushion and Richard Thomas. "From quantitative precision to qualitative judgements: Professional perspectives about the impartiality of television news during the 2015 UK General Election". In: *Journalism* 20.3 (2019), pp. 392–409.

[65] DataReportal, We Are Social, and Meltwater. *Most popular reasons for using the internet worldwide as of 3rd quarter 2023*. https://www.statista.com/statistics/1387375/internet-using-global-reasons/. 2023.

[66] Don E. Davis, Kenneth Rice, Stacey McElroy, Cirleen DeBlaere, Elise Choe, Daryl R. Van Tongeren, and Joshua N. Hook. "Distinguishing Intellectual Humility and General Humility". In: *The Journal of Positive Psychology* 11 (2016), pp. 215–224. ISSN: 1743-9760. DOI: 10.1080/17439760.2015.1048818.

[67] Samantha A. Deffler, Mark R. Leary, and Rick H. Hoyle. "Knowing What You Know: Intellectual Humility and Judgments of Recognition Memory". In: *Personality and Individual Differences* 96 (2016), pp. 255–259. ISSN: 0191-8869. DOI: 10.1016/j.paid.2016.03.016.

[68] Tim Draws, Oana Inel, Nava Tintarev, Christian Baden, and Benjamin Timmermans. "Comprehensive Viewpoint Representations for a Deeper Understanding of User Interactions With Debated Topics". en. In: *Proceedings of the 2022 ACM SIGIR Conference on Human Information Interaction and Retrieval*. CHIIR '22. New York, NY, USA: ACM, 2022, p. 11. DOI: 10.1145/3498366.3505812. URL: https://drive.google.com/file/d/1cMUzKX9QkAGfTAM8WaDKRK7y23auzNn5/view?usp=sharing.

[69] Tim Draws, Jody Liu, and Nava Tintarev. "Helping users discover perspectives: Enhancing opinion mining with joint topic models". en. In: *2020 International Conference on Data Mining Workshops (ICDMW)*. Sorrento, Italy: IEEE, Nov. 2020, pp. 23–30. ISBN: 978-1-72819-012-9. DOI: 10.1109/ICDMW51313.2020.00013. URL: https://ieeexplore.ieee.org/document/9346407/ (visited on 07/13/2021).

[70] Tim Draws, Karthikeyan Natesan Ramamurthy, Ioana Baldini, Amit Dhurandhar, Inkit Padhi, Benjamin Timmermans, and Nava Tintarev. "Explainable Cross-Topic Stance Detection for Search Results". In: *ACM SIGIR Conference on Human Information Interaction and Retrieval*. CHIIR '23. New York, NY, USA: ACM, 2023. DOI: 10.1145/3576840.3578296.

[71] Tim Draws, **Alisa Rieger**, Oana Inel, Ujwal Gadiraju, and Nava Tintarev. "A Checklist to Combat Cognitive Biases in Crowdsourcing". In: *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 9 (2021), pp. 48–59. ISSN: 2769-1349. DOI: 10.1609/hcomp.v9i1.18939.

[72] Tim Draws, Nirmal Roy, Oana Inel, **Alisa Rieger**, Rishav Hada, Mehmet Orcun Yalcin, Benjamin Timmermans, and Nava Tintarev. "Viewpoint Diversity in Search Results". In: *Advances in Information Retrieval*. Ed. by Jaap Kamps, Lorraine Goeuriot, Fabio Crestani, Maria Maistro, Hideo Joho, Brian Davis, Cathal Gurrin, Udo Kruschwitz, and Annalina Caputo. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023, pp. 279–297. ISBN: 978-3-031-28244-7. DOI: 10.1007/978-3-031-28244-7_18.

[73] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. "Assessing Viewpoint Diversity in Search Results Using Ranking Fairness Metrics". en. In: *ACM SIGKDD Explorations Newsletter* 23.1 (May 2021), pp. 50–58. ISSN: 1931-0145, 1931-0153. DOI: 10.1145/3468507.3468515. (Visited on 07/13/2021).

[74] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. "This Is Not What We Ordered: Exploring Why Biased Search Result Rankings Affect User Attitudes on Debated Topics". en. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Virtual Event Canada: ACM, July 2021, pp. 295–305. ISBN: 978-1-4503-8037-9. DOI: 10.1145/3404835.3462851. (Visited on 07/13/2021).

[75] Tim Alexander Draws. "Understanding Viewpoint Biases in Web Search Results". PhD thesis. Delft, Netherlands: Delft University of Technology, 2023. DOI: 10.4233/uuid:1b177026-6af7-48f3-ba04-ab7109db3c36.

[76] Marina Drosou and Evaggelia Pitoura. "Search result diversification". en. In: *SIGMOD Record* 39.1 (2010), p. 7.

[77] Max Z van Drunen, Natali Helberger, and Mariella Bastian. "Know your algorithm: what media organizations need to explain to their users about news personalization". In: *International Data Privacy Law* 9.4 (2019), pp. 220–235.

[78] Lorik Dumani, Patrick J Neumann, and Ralf Schenkel. "A framework for argument retrieval: Ranking argument clusters by frequency and specificity". In: *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part I*. Springer. 2020, pp. 431–445.

[79] Karl Duncker. "On Problem-Solving". In: *Psychological Monographs* 58 (1945). Ed. by Lynne S. Lees, pp. i–113. ISSN: 0096-9753. DOI: 10.1037/h0093599.

[80] Carsten Eickhoff. "Cognitive Biases in Crowdsourcing". In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. Marina Del Rey CA USA: ACM, 2018, pp. 162–170. ISBN: 978-1-4503-5581-0. DOI: 10.1145/3159652.3159654.

[81] Robert Epstein, Michael Lothringer, and Vanessa R. Zankich. "How a Daily Regimen of Operant Conditioning Might Explain the Power of the Search Engine Manipulation Effect (SEME)". In: *Behavior and Social Issues* (2024). ISSN: 2376-6786. DOI: 10.1007/s42822-023-00155-0.

[82] Robert Epstein and Ronald E. Robertson. "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections". In: *Proceedings of the National Academy of Sciences* 112 (Aug. 2015), E4512–E4521. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1419828112.

[83] Robert Epstein, Ronald E. Robertson, David Lazer, and Christo Wilson. "Suppressing the Search Engine Manipulation Effect (SEME)". en. In: *Proceedings of the ACM on Human-Computer Interaction* 1.CSCW (Dec. 2017), pp. 1–22. ISSN: 2573-0142. DOI: 10.1145/3134677. (Visited on 07/13/2021).

[84] Sarah Eskens, Natali Helberger, and Judith Moeller. "Challenged by news personalisation: five perspectives on the right to receive information". en. In: *Journal of Media Law* 9.2 (July 2017), pp. 259–284. ISSN: 1757-7632, 1757-7640. DOI: 10.1080/17577632.2017.1387353. (Visited on 07/13/2021).

[85] European Commission. Directorate General for Communications Networks, Content and Technology. *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self Assessment.* Tech. rep. LU: Publications Office, 2020.

[86] Eurostat. *Share of daily internet users in the European Union (EU-27) from 2013 to 2022.* https://www.statista.com/statistics/1238307/eu-european-union-internet-users-use-accessed-internet-daily/. 2023.

[87] Jonathan St BT Evans. "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition". In: *Annu. Rev. Psychol.* 59 (2008), pp. 255–278.

[88] Álvaro Figueira and Luciana Oliveira. "The Current State of Fake News: Challenges and Opportunities". In: *Procedia Computer Science.* CENTERIS 2017 - International Conference on ENTERprise Information Systems / ProjMAN 2017 - International Conference on Project MANagement / HCist 2017 - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2017 121 (2017), pp. 817–825. ISSN: 1877-0509. DOI: 10.1016/j.procs.2017.11.106.

[89] Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White. "Evaluating Implicit Measures to Improve Web Search". In: *ACM Transactions on Information Systems* 23 (2005), pp. 147–168. ISSN: 1046-8188. DOI: 10.1145/1059981.1059982.

[90] Shane Frederick. "Cognitive Reflection and Decision Making". In: *Journal of Economic Perspectives* 19.4 (Nov. 2005), pp. 25–42. ISSN: 0895-3309. DOI: 10.1257/089533005775196732. (Visited on 03/24/2021).

[91] Isabelle Freiling, Nicole M. Krause, and Dietram A. Scheufele. "Science and Ethics of "Curing" Misinformation". In: *AMA Journal of Ethics* 25 (2023), pp. 228–237. ISSN: 2376-6980. DOI: 10.1001/amajethics.2023.228.

[92] Katja Freistein, Frank Gadinger, and Christine Unrau. "It just feels right. Visuality and emotion norms in right-wing populist storytelling". In: *International Political Sociology* 16.4 (2022), olac017.

[93] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and Peter Holtz. "Analyzing Knowledge Gain of Users in Informational Search Sessions on the Web". In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval.* CHIIR '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 2–11. ISBN: 978-1-4503-4925-3. DOI: 10.1145/3176349.3176381.

[94] Ruoyuan Gao and Chirag Shah. "Toward Creating a Fairer Ranking in Search Engine Results". In: *Information Processing & Management* 57 (2020), p. 102138. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2019.102138.

[95] Maria Garrido and Stephen Wyber. *Development and Access to Information 2017.* Report. International Federation of Library Associations and Institutions, 2017.

[96] Lisa Gevelber. "How Mobile Has Changed How People Get Things Done: New Consumer Behavior Data". In: *Think with Google* (2016). URL: https://think.storage.googleapis.com/docs/mobile-search-consumer-behavior-data.pdf.

[97] Lisa Gevelber. *It's all about 'me'—how people are taking search personally.* Tech. rep. 2018. URL: https://www.thinkwithgoogle.com/marketing-strategies/search/personal-needs-search-trends/.

[98] Gizem Gezici, Aldo Lipani, Yucel Saygin, and Emine Yilmaz. "Evaluation metrics for measuring bias in search engine results". en. In: *Information Retrieval Journal* 24.2 (Apr. 2021), pp. 85–113. ISSN: 1386-4564, 1573-7659. DOI: 10.1007/s10791-020-09386-w. (Visited on 07/13/2021).

[99] Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. "A Think-Aloud Study to Understand Factors Affecting Online Health Search". In: *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval.* Vancouver BC Canada: ACM, Mar. 2020, pp. 273–282. ISBN: 978-1-4503-6892-6. DOI: 10.1145/3343413.3377961.

[100] Gerd Gigerenzer. "On the Supposed Evidence for Libertarian Paternalism". In: *Review of Philosophy and Psychology* 6 (2015), pp. 361–383. ISSN: 1878-5166. DOI: 10.1007/s13164-015-0248-1.

[101] Gerd Gigerenzer. "Why Heuristics Work". In: *Perspectives on Psychological Science* 3 (Jan. 2008), pp. 20–29. ISSN: 1745-6916, 1745-6924. DOI: 10.1111/j.1745-6916.2008.00058.x.

[102] Gerd Gigerenzer and Henry Brighton. "Homo heuristicus: Why biased minds make better inferences". In: *Topics in cognitive science* 1.1 (2009), pp. 107–143.

[103] Fausto Giunchiglia, Styliani Kleanthous, Jahna Otterbacher, and Tim Draws. "Transparency Paths - Documenting the Diversity of User Perceptions". en. In: *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization.* Utrecht Netherlands: ACM, June 2021, pp. 415–420. ISBN: 978-1-4503-8367-7. DOI: 10.1145/3450614.3463292. (Visited on 07/13/2021).

[104] Tim Gorichanaz. "Relating Information Seeking and Use to Intellectual Humility". In: *Journal of the Association for Information Science and Technology* 73 (2022), pp. 643–654. ISSN: 2330-1643. DOI: 10.1002/asi.24567.

[105] Tim Gorichanaz. "Virtuous Search: A Framework for Intellectual Virtue in Online Search". In: *Journal of the Association for Information Science and Technology* n/a (2023). ISSN: 2330-1643. DOI: 10.1002/asi.24832.

[106] Eduardo Graells-Garrido, Mounia Lalmas, and Ricardo Baeza-Yates. "Data Portraits and Intermediary Topics: Encouraging Exploration of Politically Diverse Profiles". In: *Proceedings of the 21st International Conference on Intelligent User Interfaces.* 2016, pp. 228–240.

[107] Mark Graham and William H. Dutton. *Society and the Internet: How Networks of Information and Communication Are Changing Our Lives.* Oxford University Press, 2019. ISBN: 978-0-19-187932-6. DOI: 10.1093/oso/9780198843498.001.0001.

[108] Tim Groot Kormelink and Irene Costera Meijer. "A user perspective on time spent: temporal experiences of everyday news use". In: *Journalism Studies* 21.2 (2020), pp. 271–286.

[109] Igor Grossmann, Mayumi Karasawa, Satoko Izumi, Jinkyung Na, Michael E. W. Varnum, Shinobu Kitayama, and Richard E. Nisbett. "Aging and Wisdom: Culture Matters". In: *Psychological Science* 23 (2012), pp. 1059–1066. ISSN: 0956-7976. DOI: 10.1177/0956797612446025.

[110] Zhiwei Guan and Edward Cutrell. "An Eye Tracking Study of the Effect of Target Rank on Web Search". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '07. New York, NY, USA: Association for Computing Machinery, 2007, pp. 417–420. ISBN: 978-1-59593-593-9. DOI: 10.1145/1240624.1240691.

[111] Andrew Guess, Jonathan Nagler, and Joshua Tucker. "Less than you think: Prevalence and predictors of fake news dissemination on Facebook". In: *Science advances* 5.1 (2019), eaau4586.

[112] Akshit Gupta, Debadeep Basu, Ramya Ghantasala, Sihang Qiu, and Ujwal Gadiraju. "To trust or not to trust: How a conversational interface affects trust in a decision support system". In: *Proceedings of the ACM Web Conference 2022.* 2022, pp. 3531–3540.

[113] Jürgen Habermas. "Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research1". In: *Communication Theory* 16 (2006), pp. 411–426. ISSN: 1050-3293. DOI: 10.1111/j.1468-2885.2006.00280.x.

[114] Jutta Haider and Olof Sundin. "Information Literacy Challenges in Digital Culture: Conflicting Engagements of Trust and Doubt". In: *Information, Communication & Society* 25 (2022), pp. 1176–1191. ISSN: 1369-118X. DOI: 10.1080/1369118X.2020.1851389.

[115] Jutta Haider and Olof Sundin. *Invisible Search and Online Search Engines: The Ubiquity of Search in Everyday Life*. Taylor & Francis, 2019. DOI: 10.4324/9780429448546.

[116] Matthew Haigh. "Has the Standard Cognitive Reflection Test Become a Victim of Its Own Success?" In: *Advances in Cognitive Psychology* 12 (2016), p. 145. DOI: 10.5709/acp-0193-5.

[117] Alexander Halavais. *Search Engine Society*. John Wiley & Sons, 2017. ISBN: 978-1-5095-1686-5.

[118] Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M Meyer, and Iryna Gurevych. "A retrospective analysis of the fake news challenge stance detection task". In: *arXiv preprint arXiv:1806.05180* (2018).

[119] Pelle Guldborg Hansen and Andreas Maaløe Jespersen. "Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy". In: *European Journal of Risk Regulation* 4.1 (2013), pp. 3–28.

[120] Jaron Harambam, Dimitrios Bountouridis, Mykola Makhortykh, and Joris Van Hoboken. "Designing for the better by taking users into account: A qualitative evaluation of user control mechanisms in (news) recommender systems". In: *Proceedings of the 13th ACM Conference on Recommender Systems*. 2019, pp. 69–77.

[121] Uriel Haran, Ilana Ritov, and Barbara A. Mellers. "The Role of Actively Open-Minded Thinking in Information Acquisition, Accuracy, and Calibration". In: *Judgment and Decision Making* 8 (2013), pp. 188–201. ISSN: 1930-2975.

[122] Eszter Hargittai, Lindsay Fullerton, Ericka Menchen-Trevino, and Kristin Yates Thomas. "Trust Online: Young Adults' Evaluation of Web Content". In: *International Journal of Communication* 4 (2010), p. 27. ISSN: 1932-8036.

[123] Sandra G. Hart and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research". en. In: *Advances in Psychology*. Ed. by Peter A. Hancock and Najmedin Meshkati. Vol. 52. Human Mental Workload. North-Holland, Jan. 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9. URL: https://www.sciencedirect.com/science/article/pii/S0166411508623869 (visited on 07/13/2022).

[124] William Hart, Dolores Albarracín, Alice H. Eagly, Inge Brechan, Matthew J. Lindberg, and Lisa Merrill. "Feeling Validated versus Being Correct: A Meta-Analysis of Selective Exposure to Information". In: *Psychological Bulletin* 135 (2009), pp. 555–588. ISSN: 1939-1455. DOI: 10.1037/a0015701.

[125] Amelia Hassoun, Ian Beacock, Sunny Consolvo, Beth Goldberg, Patrick Gage Kelley, and Daniel M. Russell. "Practicing Information Sensibility: How Gen Z Engages with Online Information". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 1–17. ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3581328.

[126] Andrew F. Hayes and Klaus Krippendorff. "Answering the Call for a Standard Reliability Measure for Coding Data". en. In: *Communication Methods and Measures* 1 (Apr. 2007), pp. 77–89. ISSN: 1931-2458, 1931-2466. DOI: 10.1080/19312450709336664.

[127] Werner Heisenberg. "Physics and Philosophy: The Revolution in Modern Science". In: *World Perspectives* 19 (1958). Ed. by Ruth Nanda Anshen.

[128] Natali Helberger. "On the Democratic Role of News Recommenders". en. In: *Digital Journalism* 7.8 (Sept. 2019), pp. 993–1012. ISSN: 2167-0811, 2167-082X. DOI: 10.1080/21670811.2019.1623700. (Visited on 07/13/2021).

[129] Natali Helberger, Kari Karppinen, and Lucia D'Acunto. "Exposure Diversity as a Design Principle for Recommender Systems". In: *Information, Communication & Society* 21 (2018), pp. 191–207. ISSN: 1369-118X. DOI: 10.1080/1369118X.2016.1271900.

[130] Natali Helberger, Katharina Kleinen-von Königslöw, and Rob van der Noll. "Regulating the New Information Intermediaries as Gatekeepers of Information Diversity". In: *info* 17 (2015). Ed. by Mr Andrea Renda and Dr Kristina Irion Dr. Luciano Morganti, pp. 50–71. ISSN: 1463-6697. DOI: 10.1108/info-05-2015-0034.

[131] Erik Hermann. "Artificial intelligence and mass personalization of communication content—An ethical and literacy perspective". In: *New media & society* 24.5 (2022), pp. 1258–1277.

[132] Ralph Hertwig and Till Grüne-Yanoff. "Nudging and Boosting: Steering or Empowering Good Decisions". In: *Perspectives on Psychological Science* 12 (Nov. 2017), pp. 973–986. ISSN: 1745-6916. DOI: 10.1177/1745691617702496.

[133] Dustin Hillard, Stefan Schroedl, Eren Manavoglu, Hema Raghavan, and Chirs Leggetter. "Improving Ad Relevance in Sponsored Search". In: *Proceedings of the Third ACM International Conference on Web Search and Data Mining*. WSDM '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 361–370. ISBN: 978-1-60558-889-6. DOI: 10.1145/1718487.1718532.

[134] Thomas T Hills. "The Dark Side of Information Proliferation". In: *Perspectives on Psychological Science* 14 (2019), pp. 323–330.

[135] Lawrence M Hinman. *Searching ethics: The role of search engines in the construction and distribution of knowledge*. Springer, 2008.

[136] Adrian Holzer, Nava Tintarev, Samuel Bendahan, Bruno Kocher, Shane Greenup, and Denis Gillet. "Digitally Scaffolding Debate in the Classroom". en. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC Canada: ACM, Apr. 2018, pp. 1–6. ISBN: 978-1-4503-5621-3. DOI: 10.1145/3170427.3188499.

[137] Marieke van Hoof, Corine S Meppelink, Judith Moeller, and Damian Trilling. "Searching differently? How political attitudes impact search queries about political issues". In: *New Media & Society* (2022), p. 14614448221104405.

[138] Marieke van Hoof, Damian Trilling, Corine Meppelink, Judith Moeller, and Felicia Loecherbach. "Googling Politics? The Computational Identification of Political and News-related Searches from Web Browser Histories". In: (2023).

[139] Steven Houben and Christian Weichel. "Overcoming Interaction Blindness through Curiosity Objects". en. In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*. Paris, France: ACM Press, 2013, p. 1539. DOI: 10.1145/2468356.2468631.

[140] Lauren C. Howe and Jon A. Krosnick. "Attitude Strength". In: *Annual Review of Psychology* 68 (2017), pp. 327–351. DOI: 10.1146/annurev-psych-122414-033600.

[141] Rick H. Hoyle, Erin K. Davisson, Kate J. Diebels, and Mark R. Leary. "Holding Specific Views with Humility: Conceptualization and Measurement of Specific Intellectual Humility". In: *Personality and Individual Differences* 97 (2016), pp. 165–172. ISSN: 0191-8869. DOI: 10.1016/j.paid.2016.03.043.

[142] Sha Hu, Zhicheng Dou, Xiaojie Wang, Tetsuya Sakai, and Ji-Rong Wen. "Search Result Diversification Based on Hierarchical Intents". en. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. Melbourne Australia: ACM, Oct. 2015, pp. 63–72. ISBN: 978-1-4503-3794-6. DOI: 10.1145/2806416.2806455. (Visited on 07/13/2021).

[143] Christoph Hube, Besnik Fetahu, and Ujwal Gadiraju. "Understanding and Mitigating Worker Biases in the Crowdsourced Collection of Subjective Judgments". en. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, May 2019, pp. 1–12. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300637.

[144] Deborah A Hwa-Froelich and Debra C Vigil. "Three aspects of cultural influence on communication: A literature review". In: *Communication Disorders Quarterly* 25.3 (2004), pp. 107–118.

[145] Peter Ingwersen and Kalervo Järvelin. "Information Retrieval in Context: IRiX". In: *SIGIR Forum* 39.2 (Dec. 2005), pp. 31–39. ISSN: 0163-5840. DOI: 10.1145/1113343.1113351.

[146] Peter Ingwersen and Kalervo Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer Science & Business Media, 2005. ISBN: 978-1-4020-3851-8.

[147] Sarah Inman and David Ribes. ""Beautiful Seams": Strategic Revelations and Concealments". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–14. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300508.

[148] Mathias Jesse and Dietmar Jannach. "Digital Nudging with Recommender Systems: Survey and Future Directions". en. In: *Computers in Human Behavior Reports* 3 (Jan. 2021), p. 100052. DOI: 10.1016/j.chbr.2020.100052.

[149] Thorsten Joachims. "Optimizing Search Engines Using Clickthrough Data". In: *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '02. New York, NY, USA: Association for Computing Machinery, 2002, pp. 133–142. ISBN: 978-1-58113-567-1. DOI: 10.1145/775047.775067.

[150] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. "Accurately Interpreting Clickthrough Data as Implicit Feedback". en. In: *ACM SIGIR Forum* 51.1 (2016), p. 8.

[151] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. "Accurately Interpreting Clickthrough Data as Implicit Feedback". In: *ACM SIGIR Forum* 51 (2017), pp. 4–11. ISSN: 0163-5840. DOI: 10.1145/3130332.3130334.

[152] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. "Unbiased Learning-to-Rank with Biased Feedback". en. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. Cambridge United Kingdom: ACM, Feb. 2017, pp. 781–789. ISBN: 978-1-4503-4675-7. DOI: 10.1145/3018661.3018699. (Visited on 07/13/2021).

[153] Dan R. Johnson, Meredith P. Murphy, and Riley M. Messer. "Reflecting on Explanatory Ability: A Mechanism for Detecting Gaps in Causal Knowledge". In: *Journal of Experimental Psychology: General* 145 (2016), pp. 573–588. ISSN: 1939-2222. DOI: 10.1037/xge0000161.

[154] Hideo Joho and Joemon M Jose. "A comparative study of the effectiveness of search result presentation on the web". In: *European Conference on Information Retrieval*. Springer. 2006, pp. 302–313.

[155] S Mo Jones-Jang, Tara Mortensen, and Jingjing Liu. "Does media literacy help identification of fake news? Information literacy helps, but other literacies don't". In: *American behavioral scientist* 65.2 (2021), pp. 371–388.

[156] Dan M. Kahan. *The Politically Motivated Reasoning Paradigm*. SSRN Scholarly Paper. Rochester, NY, 2015.

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

[158] Yvonne Kammerer and Peter Gerjets. "How search engine users evaluate and select Web search results: The impact of the search engine interface on credibility assessments". In: *Web search engine research*. Emerald Group Publishing Limited, 2012.

[159] Markus Kattenbeck and David Elsweiler. "Understanding Credibility Judgements for Web Search Snippets". In: *Aslib Journal of Information Management* 71 (2019), pp. 368–391. ISSN: 2050-3806. DOI: 10.1108/AJIM-07-2018-0181.

[160] Mesut Kaya and Derek Bridge. "Subprofile-aware diversification of recommendations". en. In: *User Modeling and User-Adapted Interaction* 29.3 (July 2019), pp. 661–700. ISSN: 0924-1868, 1573-1391. DOI: 10.1007/s11257-019-09235-6. (Visited on 07/13/2021).

[161] David Kaye. *Report on Artificial Intelligence Technologies and Implications for the Information Environment*. Tech. rep. A/73/348. Special Rapporteur on the promotion and protection of the right to …, 2018.

[162] Varol Kayhan. “Confirmation Bias: Roles of Search Engines and Search Contexts”. en. In: (2015), p. 18.

[163] Dominique Kelly, Yimin Chen, Sarah E. Cornwell, Nicole S. Delellis, Alex Mayhew, Sodiq Onaolapo, and Victoria L. Rubin. “Bing Chat: The Future of Search Engines?” In: *Proceedings of the Association for Information Science and Technology* 60 (2023), pp. 1007–1009. ISSN: 2373-9231. DOI: 10.1002/pra2.927.

[164] Matthew Kelly. “Epistemology, Epistemic Belief, Personal Epistemology, and Epistemics: A Review of Concepts as They Impact Information Behavior Research”. In: *Journal of the Association for Information Science and Technology* 72 (2021), pp. 507–519. ISSN: 2330-1643. DOI: 10.1002/asi.24422.

[165] Ann-Marie Kennedy, Katharine Jones, and Janine Williams. “Children as vulnerable consumers in online environments”. In: *Journal of Consumer Affairs* 53.4 (2019), pp. 1478–1506.

[166] Moaiad Ahmad Khder. “Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application.” In: *International Journal of Advances in Soft Computing & Its Applications* 13.3 (2021).

[167] Dam Hee Kim and Josh Pasek. “Explaining the diversity deficit: Value-trait consistency in news exposure and democratic citizenship”. In: *Communication research* 47.1 (2020), pp. 29–54.

[168] Jaewon Kim, Paul Thomas, Ramesh Sankaranarayana, Tom Gedeon, and Hwan-Jin Yoon. “Understanding eye movements on mobile devices for better presentation of search results”. In: *Journal of the Association for Information Science and Technology* 67.11 (2016), pp. 2607–2619.

[169] Jinyoung Kim, Brenna McNally, Leyla Norooz, and Allison Druin. “Internet Search Roles of Adults in Their Homes”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 4948–4959. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025572.

[170] Jan Kirchner and Christian Reuter. “Countering Fake News: A Comparison of Possible Solutions Regarding User Acceptance and Effectiveness”. en. In: *Proceedings of the ACM on Human-Computer Interaction* 4 (Oct. 2020), pp. 1–27. ISSN: 2573-0142. DOI: 10.1145/3415211.

[171] Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. “The Future of Crowd Work”. In: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. CSCW ’13. New York, NY, USA: Association for Computing Machinery, 2013, pp. 1301–1318. ISBN: 978-1-4503-1331-5. DOI: 10.1145/2441776.2441923.

[172] Silvia Knobloch-Westerwick and Jingbo Meng. “Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information”. In: *Communication Research* 36 (2009), pp. 426–448. ISSN: 0093-6502, 1552-3810. DOI: 10.1177/0093650209333030.

[173] Silvia Knobloch-Westerwick, Benjamin K. Johnson, and Axel Westerwick. “Confirmation Bias in Online Searches: Impacts of Selective Exposure Before an Election on Political Attitude Strength and Shifts”. In: *Journal of Computer-Mediated Communication* 20 (2015), pp. 171–187. ISSN: 1083-6101. DOI: 10.1111/jcc4.12105.

[174] Ansgar Koene, Elvira Perez, Christopher James Carter, Ramona Statache, Svenja Adolphs, Claire O’Malley, Tom Rodden, and Derek McAuley. “Ethics of personalized information filtering”. In: *Internet Science: Second International Conference, INSCI 2015, Brussels, Belgium, May 27-29, 2015, Proceedings 2*. Springer. 2015, pp. 123–132.

[175] Jonah Koetke, Karina Schumann, and Tenelle Porter. “Intellectual Humility Predicts Scrutiny of COVID-19 Misinformation”. In: *Social Psychological and Personality Science* 13 (2022), pp. 277–284. ISSN: 1948-5506. DOI: 10.1177/1948550620988242.

[176] Tim Groot Kormelink and Irene Costera Meijer. “What clicks actually mean: Exploring digital news user practices”. In: *Journalism* 19.5 (2018), pp. 668–683.

[177] Hilary Kornblith. “Justified Belief and Epistemically Responsible Action”. In: *The Philosophical Review* 92 (1983), pp. 33–48. ISSN: 0031-8108. DOI: 10.2307/2184520.

[178] Anastasia Kozyreva, Stephan Lewandowsky, and Ralph Hertwig. "Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools". In: *Psychological Science in the Public Interest* 21 (2020), pp. 103–156. ISSN: 1529-1006. DOI: 10.1177/1529100620946707.

[179] Anastasia Kozyreva, Philipp Lorenz-Spreen, Stefan Herzog, Ullrich Ecker, Stephan Lewandowsky, and Ralph Hertwig. "Toolbox of Interventions against Online Misinformation and Manipulation". In: (2022).

[180] Anastasia Kozyreva, Sam Wineburg, Stephan Lewandowsky, and Ralph Hertwig. "Critical Ignoring as a Core Competence for Digital Citizens". In: *Current Directions in Psychological Science* 32 (2023), pp. 81–88. ISSN: 0963-7214. DOI: 10.1177/09637214221121570.

[181] Ethan Kross and Igor Grossmann. "Boosting Wisdom: Distance from the Self Enhances Wise Reasoning, Attitudes, and Behavior". In: *Journal of Experimental Psychology: General* 141 (2012), pp. 43–48. ISSN: 1939-2222. DOI: 10.1037/a0024158.

[182] Elizabeth J. Krumrei-Mancuso, Megan C. Haggard, Jordan P. LaBouff, and Wade C. Rowatt. "Links between Intellectual Humility and Acquiring Knowledge". In: *The Journal of Positive Psychology* 15 (2020), pp. 155–170. ISSN: 1743-9760. DOI: 10.1080/17439760.2019.1579359.

[183] Elizabeth J. Krumrei-Mancuso and Brian Newman. "Intellectual Humility in the Sociopolitical Domain". In: *Self and Identity* 19 (2020), pp. 989–1016. ISSN: 1529-8868. DOI: 10.1080/15298868.2020.1714711.

[184] Dilek Küçük and Fazli Can. "Stance Detection: A Survey". en. In: *ACM Computing Surveys* 53 (May 2020), pp. 1–37. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3369026.

[185] Carol C. Kuhlthau. "Inside the Search Process: Information Seeking fcarrollrom the User's Perspective". In: *Journal of the American Society for Information Science* 42 (1991), pp. 361–371. ISSN: 1097-4571. DOI: 10.1002/(SICI)1097-4571(199106)42:5<361::AID-ASI6>3.0.CO;2-#.

[186] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi, and Karrie Karahalios. "Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media". en. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. Portland Oregon USA: ACM, Feb. 2017, pp. 417–432. ISBN: 978-1-4503-4335-0. DOI: 10.1145/2998181.2998321. (Visited on 07/13/2021).

[187] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi, and Karrie Karahalios. "Search bias quantification: investigating political bias in social media and web search". en. In: *Information Retrieval Journal* 22.1-2 (Apr. 2019), pp. 188–227. ISSN: 1386-4564, 1573-7659. DOI: 10.1007/s10791-018-9341-2. (Visited on 07/13/2021).

[188] Haewoon Kwak, Jisun An, Joni Salminen, Soon-Gyo Jung, and Bernard J. Jansen. "What We Read, What We Search: Media Attention and Public Attention Among 193 Countries". In: *Proceedings of the 2018 World Wide Web Conference*. WWW '18. Lyon, France: International World Wide Web Conferences Steering Committee, 2018, pp. 893–902. ISBN: 9781450356398. DOI: 10.1145/3178876.3186137.

[189] Monica Landoni, Mohammad Aliannejadi, Theo Huibers, Emiliana Murgia, and Maria Soledad Pera. "Have a clue! the effect of visual cues on children's search behavior in the classroom". In: *ACM SIGIR Conference on Human Information Interaction and Retrieval*. 2022, pp. 310–314.

[190] Monica Landoni, Theo Huibers, Mohammad Aliannejadi, Emiliana Murgia, and Maria Soledad Pera. "Getting to Know You: Search Logs and Expert Grading to Define Children's Search Roles in the Classroom." In: *DESIRES*. 2021, pp. 44–52.

[191] John Lawrence and Chris Reed. "Argument mining: A survey". In: *Computational Linguistics* 45.4 (2020), pp. 765–818.

[192] John Lawrence and Chris Reed. "Combining Argument Mining Techniques". en. In: *Proceedings of the 2nd Workshop on Argumentation Mining*. Denver, CO: Association for Computational Linguistics, 2015, pp. 127–136. DOI: 10.3115/v1/W15-0516. URL: http://aclweb.org/anthology/W15-0516 (visited on 07/13/2021).

[193] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[194] Huyen Le, Raven Maragh, Brian Ekdale, Andrew High, Timothy Havens, and Zubair Shafiq. "Measuring Political Personalization of Google News Search". In: *The World Wide Web Conference*. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 2957–2963. ISBN: 978-1-4503-6674-8. DOI: 10.1145/3308558.3313682.

[195] Mark R. Leary, Kate J. Diebels, Erin K. Davisson, Katrina P. Jongman-Sereno, Jennifer C. Isherwood, Kaitlin T. Raimi, Samantha A. Deffler, and Rick H. Hoyle. "Cognitive and Interpersonal Features of Intellectual Humility". In: *Personality and Social Psychology Bulletin* 43 (2017), pp. 793–813. ISSN: 0146-1672. DOI: 10.1177/0146167217697695.

[196] Neil Levy. "Nudges in a Post-Truth World". In: *Journal of medical ethics* 43 (2017), pp. 495–500.

[197] Dirk Lewandowski. "Credibility in Web Search Engines". In: *Online Credibility and Digital Ethos: Evaluating Computer-Mediated Communication*. IGI Global, 2013, pp. 131–146. ISBN: 978-1-4666-2663-8. DOI: 10.4018/978-1-4666-2663-8.ch008.

[198] Stephan Lewandowsky, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. "Misinformation and Its Correction: Continued Influence and Successful Debiasing". In: *Psychological Science in the Public Interest* 13.3 (Dec. 2012), pp. 106–131. ISSN: 1529-1006, 1539-6053. DOI: 10.1177/1529100612451018. (Visited on 03/30/2021).

[199] Stephan Lewandowsky, Laura Smillie, David Garcia, Ralph Hertwig, Jim Weatherall, Stefanie Egidy, Ronald E Robertson, Cailin O'Connor, Anastasia Kozyreva, Philipp Lorenz-Spreen, et al. "Technology and democracy: Understanding the influence of online technologies on political behaviour and decision-making". In: (2020).

[200] Stephan Lewandowsky and Sander van der Linden. "Countering Misinformation and Fake News Through Inoculation and Prebunking". In: *European Review of Social Psychology* 32 (2021), pp. 348–384. ISSN: 1046-3283. DOI: 10.1080/10463283.2021.1876983.

[201] Q. Vera Liao and Wai-Tat Fu. "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. New York, NY, USA: Association for Computing Machinery, 2014, pp. 184–196. ISBN: 978-1-4503-2540-0. DOI: 10.1145/2531602.2531711.

[202] Scott O Lilienfeld, Rachel Ammirati, and Kristin Landfield. "Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare?" In: *Perspectives on psychological science* 4 (2009), pp. 390–398.

[203] Marco Lippi and Paolo Torroni. "Context-independent claim detection for argument mining". In: *Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015.

[204] Enrico Liscio. "Axies: Identifying and Evaluating Context-Specific Values". en. In: *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems*. AAMAS '21. 2021, p. 10.

[205] Jingjing Liu and Xiangmin Zhang. "The Role of Domain Knowledge in Document Selection from Search Results". In: *Journal of the Association for Information Science and Technology* 70 (2019), pp. 1236–1247. ISSN: 2330-1643. DOI: 10.1002/asi.24199.

[206] Jiqun Liu, Matthew Mitsui, Nicholas J. Belkin, and Chirag Shah. "Task, Information Seeking Intentions, and User Behavior: Toward A Multi-level Understanding of Web Search". In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. CHIIR '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 123–132. ISBN: 978-1-4503-6025-8. DOI: 10.1145/3295750.3298922.

[207] Felicia Loecherbach, Judith Moeller, Damian Trilling, and Wouter van Atteveldt. "The unified framework of media diversity: A systematic literature review". In: *Digital Journalism* 8.5 (2020), pp. 605–642.

[208] Eugène Loos, Loredana Ivan, and Donald Leu. ""Save the Pacific Northwest tree octopus": a hoax revisited. Or: How vulnerable are school children to fake news?" In: *Information and Learning Science* (2018).

[209] Philipp Lorenz-Spreen, Michael Geers, Thorsten Pachur, Ralph Hertwig, Stephan Lewandowsky, and Stefan M. Herzog. "Boosting People's Ability to Detect Microtargeted Advertising". In: *Scientific Reports* 11 (July 2021), p. 15541. ISSN: 2045-2322. DOI: 10.1038/s41598-021-94796-z.

[210] Philipp Lorenz-Spreen, Stephan Lewandowsky, Cass R. Sunstein, and Ralph Hertwig. "How Behavioural Sciences Can Promote Truth, Autonomy and Democratic Discourse Online". In: *Nature Human Behaviour* 4 (Nov. 2020), pp. 1102–1109. ISSN: 2397-3374. DOI: 10.1038/s41562-020-0889-7.

[211] Philipp Lorenz-Spreen, Lisa Oswald, Stephan Lewandowsky, and Ralph Hertwig. "A Systematic Review of Worldwide Causal and Correlational Evidence on Digital Media and Democracy". In: *Nature Human Behaviour* 7 (2023), pp. 74–101. ISSN: 2397-3374. DOI: 10.1038/s41562-022-01460-1.

[212] Ramona Ludolph, Ahmed Allam, and Peter J Schulz. "Manipulating Google's Knowledge Graph Box to Counter Biased Information Processing During an Online Search on Vaccination: Application of a Technological Debiasing Strategy". en. In: *Journal of Medical Internet Research* 18.6 (June 2016), e137. ISSN: 1438-8871. DOI: 10.2196/jmir.5430. URL: http://www.jmir.org/2016/6/e137/ (visited on 07/13/2021).

[213] Emma Lurie and Eni Mustafaraj. "Investigating the Effects of Google's Search Engine Result Page in Evaluating the Credibility of Online News Sources". In: *Proceedings of the 10th ACM Conference on Web Science*. WebSci '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 107–116. ISBN: 978-1-4503-5563-6. DOI: 10.1145/3201064.3201095.

[214] Astrid Mager. "Algorithmic Ideology". In: *Information, Communication & Society* 15 (2012), pp. 769–787. ISSN: 1369-118X. DOI: 10.1080/1369118X.2012.676056.

[215] Gary Marchionini. "Exploratory search: from finding to understanding". In: *Communications of the ACM* 49.4 (2006), pp. 41–46.

[216] Nicolas Mattis, Philipp Masur, Judith Möller, and Wouter van Atteveldt. "Nudging towards news diversity: A theoretical framework for facilitating diverse news consumption through recommender design". In: *new media & society* (2022), p. 14614448221104413.

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

[218] Daniel McDuff, Paul Thomas, Nick Craswell, Kael Rowan, and Mary Czerwinski. "Do Affective Cues Validate Behavioural Metrics for Search?" In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2021, pp. 1544–1553.

[219] Dana Mckay, Stephann Makri, Marisela Gutierrez-Lopez, Andrew MacFarlane, Sondess Missaoui, Colin Porlezza, and Glenda Cooper. "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*. CHIIR '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 173–182. ISBN: 978-1-4503-6892-6. DOI: 10.1145/3343413.3377975.

[220] Edgar Meij, Marc Bron, Laura Hollink, Bouke Huurnink, and Maarten de Rijke. "Learning semantic query suggestions". In: *International Semantic Web Conference*. Springer. 2009, pp. 424–440.

[221] Paul Mena. "Cleaning Up Social Media: The Effect of Warning Labels on Likelihood of Sharing False News on Facebook". In: *Policy & Internet* 12.2 (2020), pp. 165–183. ISSN: 1944-2866. DOI: 10.1002/poi3.214. (Visited on 03/26/2021).

[222] Ethan A. Meyers, Martin H. Turpin, Michał Białek, Jonathan A. Fugelsang, and Derek J. Koehler. "Inducing Feelings of Ignorance Makes People More Receptive to Expert (Economist) Opinion". In: *Judgment and Decision Making* 15 (2020), pp. 909–925. ISSN: 1930-2975. DOI: 10.1017/S1930297500008135.

[223] Microsoft. *Web Search API: Microsoft Bing*. 2021. URL: https://www.microsoft.com/en-us/bing/apis/bing-web-search-api.

[224] Martijn Millecamp, Robin Haveneers, and Katrien Verbert. "Cogito Ergo Quid? The Effect of Cognitive Style in a Transparent Mobile Music Recommender System". In: *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '20: 28th ACM Conference on User Modeling, Adaptation and Personalization. Genoa Italy: ACM, July 2020, pp. 323–327. ISBN: 978-1-4503-6861-2. DOI: 10.1145/3340631.3394871. (Visited on 04/03/2021).

[225] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. "To Explain or Not to Explain: The Effects of Personal Characteristics When Explaining Music Recommendations". In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*. IUI '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 397–407. ISBN: 978-1-4503-6272-6. DOI: 10.1145/3301275.3302313.

[226] Boaz Miller and Isaac Record. "Justified Belief in the Digital Age: On the Epistemic Implications of Secret Internet Technologies". In: *Episteme* 10 (2013), pp. 117–134. ISSN: 1742-3600, 1750-0117. DOI: 10.1017/epi.2013.11.

[227] Henry Milner. *Civic literacy: How informed citizens make democracy work*. UPNE, 2002.

[228] Ashlee Milton and Maria Soledad Pera. "Into the Unknown: Exploration of Search Engines' Responses to Users with Depression and Anxiety". In: *ACM Transactions on the Web* (2021).

[229] Julia A. Minson, Frances S. Chen, and Catherine H. Tinsley. "Why Won't You Listen to Me? Measuring Receptiveness to Opposing Views". In: *Management Science* (2019). DOI: 10.1287/mnsc.2019.3362.

[230] Fauzan Misra. "Accountability pressure as debiaser for confirmation bias in information search and tax consultant's recommendations". en. In: *Journal of Indonesian Economy and Business* 34 (July 2019), p. 80. ISSN: 2338-5847, 2085-8272. DOI: 10.22146/jieb.40019.

[231] Saif M. Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. "Semeval-2016 Task 6: Detecting Stance in Tweets". In: *Proceedings of the International Workshop on Semantic Evaluation*. SemEval '16. San Diego, California, June 2016.

[232] Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. "Stance and Sentiment in Tweets". In: *Special Section of the ACM Transactions on Internet Technology on Argumentation in Social Media* 17.3 (2017).

[233] Angela Molem, Stephann Makri, and Dana Mckay. "Keepin' It Reel: Investigating How Short Videos on TikTok and Instagram Reels Influence View Change". In: *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*. CHIIR '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 317–327. DOI: 10.1145/3627508.3638341.

[234] Robert J Moore and Raphael Arar. *Conversational UX design: A practitioner's guide to the natural conversation framework*. Morgan & Claypool, 2019.

[235] Mohsen Mosleh, Gordon Pennycook, Antonio A. Arechar, and David G. Rand. "Cognitive Reflection Correlates with Behavior on Twitter". In: *Nature Communications* 12 (2021), p. 921. ISSN: 2041-1723. DOI: 10.1038/s41467-020-20043-0.

[236] Abbe Mowshowitz and Akira Kawaguchi. "Assessing bias in search engines". In: *Information Processing & Management* 38.1 (2002), pp. 141–156.

[237] Jonathan Mummolo. "News from the other side: How topic relevance limits the prevalence of partisan selective exposure". In: *The Journal of Politics* 78.3 (2016), pp. 763–773.

[238] Sean A Munson, Stephanie Y Lee, and Paul Resnick. "Encouraging Reading of Diverse Political Viewpoints with a Browser Widget". In: *Seventh International Aaai Conference on Weblogs and Social Media*. 2013.

[239] Sean A Munson and Paul Resnick. "Presenting Diverse Political Opinions: How and How Much". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2010, pp. 1457–1466.

[240] N.D. *Global Charter of Ethics for Journalists*. https://www.ifj.org/who/rules-and-policy/global-charter-of-ethics-for-journalists. 2019.

[241] N.D. *Introducing chatgpt*. 2023. URL: https://openai.com/blog/chatgpt/.

[242] Philip M. Napoli. “Deconstructing the diversity principle”. In: *Journal of Communication* 49.4 (1999), pp. 7–34. ISSN: 0021-9916. DOI: 10.1111/j.1460-2466.1999.tb02815.x.

[243] Julia Neidhardt, Hannes Werthner, and Stefan Woltran. “It Is Simple, It Is Complicated”. In: *Perspectives on Digital Humanism* (2022), p. 335.

[244] Raymond S Nickerson. “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises”. en. In: (1998), p. 46.

[245] Vlad Niculae, Joonsuk Park, and Claire Cardie. “Argument Mining with Structured SVMs and RNNs”. en. In: (Apr. 2017). arXiv: 1704.06869 [cs].

[246] Safiya Umoja Noble. “Algorithms of Oppression: How Search Engines Reinforce Racism”. In: *Algorithms of Oppression*. New York University Press, 2018. ISBN: 978-1-4798-3364-1. DOI: 10.18574/nyu/9781479833641.001.0001.

[247] Alamir Novin and Eric Meyers. “Making Sense of Conflicting Science Information: Exploring Bias in the Search Engine Result Page”. en. In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. Oslo Norway: ACM, Mar. 2017, pp. 175–184. ISBN: 978-1-4503-4677-1. DOI: 10.1145/3020165.3020185. (Visited on 07/13/2021).

[248] Helga Nowotny. “Digital Humanism: Navigating the Tensions Ahead”. In: *Perspectives on Digital Humanism* (2022), p. 317.

[249] Brendan Nyhan and Jason Reifler. “When corrections fail: The persistence of political misperceptions”. In: *Political Behavior* 32.2 (2010), pp. 303–330.

[250] Bruno Oliveira and Carla Teixeira Lopes. “The Evolution of Web Search User Interfaces - An Archaeological Analysis of Google Search Engine Result Pages”. In: *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. CHIIR ’23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 55–68. DOI: 10.1145/3576840.3578320.

[251] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. “Social data: Biases, methodological pitfalls, and ethical boundaries”. In: *Frontiers in big data* 2 (2019), p. 13.

[252] Anna-Marie Ortloff, Steven Zimmerman, David Elsweiler, and Niels Henze. “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*. CHIIR ’21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 63–73. ISBN: 978-1-4503-8055-3. DOI: 10.1145/3406522.3446014.

[253] Margit E Oswald and Stefan Grosjean. “Confirmation bias”. In: *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* 79 (2004), p. 83.

[254] Jahna Otterbacher. “Addressing social bias in information retrieval”. In: *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer. 2018, pp. 121–127.

[255] Jahna Otterbacher, Jo Bates, and Paul Clough. “Competent Men and Warm Women: Gender Stereotypes and Backlash in Image Search Results”. en. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Denver Colorado USA: ACM, May 2017, pp. 6620–6631. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025727. (Visited on 07/13/2021).

[256] Jahna Otterbacher, Alessandro Checco, Gianluca Demartini, and Paul Clough. “Investigating User Perception of Gender Bias in Image Search: The Role of Sexism”. en. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. Ann Arbor MI USA: ACM, June 2018, pp. 933–936. ISBN: 978-1-4503-5657-2. DOI: 10.1145/3209978.3210094. (Visited on 07/13/2021).

[257] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. “In Google We Trust: Users’ Decisions on Rank, Position, and Relevance”. en. In: *Journal of Computer-Mediated Communication* 12.3 (Apr. 2007), pp. 801–823. ISSN: 10836101, 10836101. DOI: 10.1111/j.1083-6101.2007.00351.x. URL: https://academic.oup.com/jcmc/article/12/3/801-823/4582975 (visited on 07/13/2021).

[258] Eli Pariser. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin UK, 2011.

[259] Sachin Pathiyan Cherumanal, Damiano Spina, Falk Scholer, and W Bruce Croft. “Evaluating fairness in argument retrieval”. In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2021, pp. 3363–3367.

[260] Rik Peels. *Responsible Belief: A Theory in Ethics and Epistemology*. Oxford University Press, 2016. ISBN: 978-0-19-060812-5.

[261] Gordon Pennycook and David G. Rand. “Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning than by Motivated Reasoning”. In: *Cognition*. The Cognitive Science of Political Thought 188 (July 2019), pp. 39–50. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2018.06.011.

[262] Richard E. Petty and John T. Cacioppo. “The Elaboration Likelihood Model of Persuasion”. In: *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. Springer Series in Social Psychology. New York, NY: Springer, 1986, pp. 1–24. ISBN: 978-1-4612-4964-1. DOI: 10.1007/978-1-4612-4964-1_1.

[263] Gloria Phillips-Wren and Monica Adya. “Decision Making under Stress: The Role of Information Overload, Time Pressure, Complexity, and Uncertainty”. In: *Journal of Decision Systems* 29 (Aug. 2020), pp. 213–225. ISSN: 1246-0125. DOI: 10.1080/12460125.2020.1768680.

[264] Gloria Phillips-Wren, Theresa Jefferson, and Sueanne McKniff. “Cognitive bias and decision aid use under stressful conditions”. In: *Journal of Decision Systems* 28.2 (2019), pp. 162–184.

[265] Ari Pirkola. “The Effectiveness of Web Search Engines to Index New Sites from Different Countries.” In: *Information Research: An International Electronic Journal* 14.2 (2009).

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

[267] Tenelle Porter, Abdo Elnakouri, Ethan A. Meyers, Takuya Shibayama, Eranda Jayawickreme, and Igor Grossmann. “Predictors and Consequences of Intellectual Humility”. In: *Nature Reviews Psychology* 1 (2022), pp. 524–536. ISSN: 2731-0574. DOI: 10.1038/s44159-022-00081-9.

[268] Tenelle Porter and Karina Schumann. “Intellectual Humility and Openness to the Opposing View”. In: *Self and Identity* 17 (2018), pp. 139–162. ISSN: 1529-8868. DOI: 10.1080/15298868.2017.1361861.

[269] Tenelle Porter, Karina Schumann, Diana Selmeczy, and Kali Trzesniewski. “Intellectual Humility Predicts Mastery Behaviors When Learning”. In: *Learning and Individual Differences* 80 (2020), p. 101888. ISSN: 1041-6080. DOI: 10.1016/j.lindif.2020.101888.

[270] Suppanut Pothirattanachaikul, Takehiro Yamamoto, Yusuke Yamamoto, and Masatoshi Yoshikawa. “Analyzing the Effects of "People Also Ask" on Search Behaviors and Beliefs”. In: *Proceedings of the 31st ACM Conference on Hypertext and Social Media*. HT ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 101–110. ISBN: 978-1-4503-7098-1. DOI: 10.1145/3372923.3404786.

[271] Emmanuel M. Pothos, Stephan Lewandowsky, Irina Basieva, Albert Barque-Duran, Katy Tapper, and Andrei Khrennikov. “Information Overload for (Bounded) Rational Agents”. In: *Proceedings of the Royal Society B: Biological Sciences* 288 (Feb. 2021), p. 20202957. DOI: 10.1098/rspb.2020.2957.

[272] Martin Potthast, Lukas Gienapp, Florian Euchner, Nick Heilenkötter, Nico Weidmann, Henning Wachsmuth, Benno Stein, and Matthias Hagen. “Argument search: Assessing argument relevance”. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2019, pp. 1117–1120.

[273] Ronak Pradeep, Xueguang Ma, Rodrigo Nogueira, and Jimmy Lin. “Vera: Prediction Techniques for Reducing Harmful Misinformation in Consumer Health Search”. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 2066–2070. ISBN: 978-1-4503-8037-9. DOI: 10.1145/3404835.3463120.

[274] Erich Prem. “Our Digital Mirror”. In: *Perspectives on Digital Humanism* (2022), p. 89.

[275] ProCon.org. *Homepage*. May 2021. URL: https://www.procon.org/.

[276] Prolific. *Prolific*. Jan. 2023. URL: https://www.prolific.co/.

[277] Prolific. *Quickly find research participants you can trust*. 2021. URL: https://www.prolific.co/.

[278] Kristen Purcell, Lee Rainie, and Joanna Brenner. “Search engine use 2012”. In: (2012).

[279] Cornelius Puschmann. “Beyond the Bubble: Assessing the Diversity of Political Search Results”. In: *Digital Journalism* 7 (2019), pp. 824–843. ISSN: 2167-0811. DOI: 10.1080/21670811.2018.1539626.

[280] Minghui Qiu and Jing Jiang. “A Latent Variable Model for Viewpoint Discovery from Threaded Forum Posts”. en. In: *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies*. Association for Computational Linguistics, 2013, pp. 1031–1040. URL: https://ink.library.smu.edu.sg/sis_research/1890/.

[281] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. “Improving worker engagement through conversational microtask crowdsourcing”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–12.

[282] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. “Towards memorable information retrieval”. In: *Proceedings of the 2020 ACM SIGIR on international conference on theory of information retrieval*. 2020, pp. 69–76.

[283] Qualtrics. *Qualtrics XM - Experience Management Software*. May 2021. URL: https://www.qualtrics.com/.

[284] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2020. URL: https://www.R-project.org/.

[285] Filip Radlinski and Nick Craswell. “A Theoretical Framework for Conversational Search”. In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. CHIIR ’17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 117–126. ISBN: 978-1-4503-4677-1. DOI: 10.1145/3020165.3020183.

[286] Jerome Ramos and Carsten Eickhoff. “Search result explanations improve efficiency and trust”. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2020, pp. 1597–1600.

[287] Samuli Reijula and Ralph Hertwig. “Self-Nudging and the Citizen Choice Architect”. In: *Behavioural Public Policy* 6 (2022), pp. 119–149. ISSN: 2398-063X, 2398-0648. DOI: 10.1017/bpp.2020.5.

[288] Pengjie Ren, Zhumin Chen, Zhaochun Ren, Evangelos Kanoulas, Christof Monz, and Maarten De Rijke. “Conversations with Search Engines: SERP-based Conversational Response Generation”. In: *ACM Transactions on Information Systems* 39 (2021), 47:1–47:29. ISSN: 1046-8188. DOI: 10.1145/3432726.

[289] Arno J. Rethans, John L. Swasy, and Lawrence J. Marks. “Effects of Television Commercial Repetition, Receiver Knowledge, and Commercial Length: A Test of the Two-Factor Model”. en. In: *Journal of Marketing Research* 23 (Feb. 1986), pp. 50–61. ISSN: 0022-2437. DOI: 10.1177/002224378602300106.

[290] Myrthe Reuver, Nicolas Mattis, Marijn Sax, Suzan Verberne, Nava Tintarev, Natali Helberger, Judith Moeller, Sanne Vrijenhoek, Antske Fokkens, and Wouter van Atteveldt. “Are we human, or are we users? The role of natural language processing in human-centric news recommenders that nudge users to diverse content”. In: *Proceedings of the 1st Workshop on NLP for Positive Impact*. 2021, pp. 47–59.

[291] Rezvaneh Rezapour, Ly Dinh, and Jana Diesner. "Incorporating the Measurement of Moral Foundations Theory into Analyzing Stances on Controversial Topics". en. In: *Proceedings of the 32st ACM Conference on Hypertext and Social Media*. Virtual Event USA: ACM, Aug. 2021, pp. 177–188. ISBN: 978-1-4503-8551-0. DOI: 10.1145/3465336.3475112. (Visited on 09/06/2021).

[292] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "" Why should i trust you?" Explaining the predictions of any classifier". In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016, pp. 1135–1144.

[293] **Alisa Rieger**. "Interactive Interventions to Mitigate Cognitive Bias". In: *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 316–320. ISBN: 978-1-4503-9207-5. DOI: 10.1145/3503252.3534362.

[294] **Alisa Rieger**, Frank Bredius, Mariët Theune, and Maria Soledad Pera. "From Potential to Practice: Intellectual Humility During Search on Debated Topics". In: *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*. CHIIR '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 130–141. DOI: 10.1145/3627508.3638306.

[295] **Alisa Rieger**, Frank Bredius, Nava Tintarev, and Maria Soledad Pera. "Searching for the Whole Truth: Harnessing the Power of Intellectual Humility to Boost Better Search on Debated Topics". In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI EA '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 1–8. ISBN: 978-1-4503-9422-2. DOI: 10.1145/3544549.3585693.

[296] **Alisa Rieger**, Tim Draws, Nicolas Mattis, David Maxwell, David Elsweiler, Ujwal Gadiraju, Dana McKay, Alessandro Bozzon, and Maria Soledad Pera. "Responsible Opinion Formation on Debated Topics in Web Search". In: *Advances in Information Retrieval*. Ed. by Nazli Goharian, Nicola Tonellotto, Yulan He, Aldo Lipani, Graham McDonald, Craig Macdonald, and Iadh Ounis. Cham: Springer Nature Switzerland, 2024, pp. 437–465. ISBN: 978-3-031-56066-8. DOI: 10.1007/978-3-031-56066-8_32.

[297] **Alisa Rieger**, Tim Draws, Mariët Theune, and Nava Tintarev. "Nudges to Mitigate Confirmation Bias during Web Search on Debated Topics: Support vs. Manipulation". In: *ACM Transactions on the Web* (2023). DOI: 10.1145/3635034.

[298] **Alisa Rieger**, Tim Draws, Mariët Theune, and Nava Tintarev. "This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias". In: *Proceedings of the 32st ACM Conference on Hypertext and Social Media*. Virtual Event USA: ACM, Aug. 2021, pp. 189–199. ISBN: 978-1-4503-8551-0. DOI: 10.1145/3465336.3475101.

[299] **Alisa Rieger**, Suleiman Kulane, Ujwal Gadiraju, and Maria Soledad Pera. "Disentangling Web Search on Debated Topics: A User-Centered Exploration". In: *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 24–35. ISBN: 9798400704338. DOI: 10.1145/3627043.3659559.

[300] **Alisa Rieger**, Qurat-Ul-Ain Shaheen, Carles Sierra, Mariet Theune, and Nava Tintarev. "Towards Healthy Engagement with Online Debates: An Investigation of Debate Summaries and Personalized Persuasive Suggestions". In: *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '22 Adjunct. New York, NY, USA: Association for Computing Machinery, 2022, pp. 192–199. ISBN: 978-1-4503-9232-7. DOI: 10.1145/3511047.3537692.

[301] **Alisa Rieger**, Mariët Theune, and Nava Tintarev. "Toward Natural Language Mitigation Strategies for Cognitive Biases in Recommender Systems". In: *2nd Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence*. Dublin, Ireland: Association for Computational Linguistics, Nov. 2020, pp. 50–54.

[302] Ronald E. Robertson, Jon Green, Damian J. Ruck, Katherine Ognyanova, Christo Wilson, and David Lazer. "Users Choose to Engage with More Partisan News than They Are Exposed to on Google Search". In: *Nature* (2023), pp. 1–7. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06078-5.

[303] Ronald E. Robertson, David Lazer, and Christo Wilson. "Auditing the Personalization and Composition of Politically-Related Search Engine Results Pages". In: *Proceedings of the 2018 World Wide Web Conference*. WWW '18. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018, pp. 955–965. ISBN: 978-1-4503-5639-8. DOI: 10.1145/3178876.3186143.

[304] Jon Roozenbeek and Sander van der Linden. "Fake News Game Confers Psychological Resistance against Online Misinformation". In: *Palgrave Communications* 5 (2019), pp. 1–10. ISSN: 2055-1045. DOI: 10.1057/s41599-019-0279-9.

[305] Shawn Rosenberg. "Democracy Devouring Itself: The Rise of the Incompetent Citizen and the Appeal of Right Wing Populism". In: (2023).

[306] Nirmal Roy, Manuel Valle Torre, Ujwal Gadiraju, David Maxwell, and Claudia Hauff. "Note the highlight: incorporating active reading tools in a search as learning environment". In: *Proceedings of the 2021 conference on human information interaction and retrieval*. 2021, pp. 229–238.

[307] Tuukka Ruotsalo, Giulio Jacucci, Petri Myllymäki, and Samuel Kaski. "Interactive Intent Modeling: Information Discovery beyond Search". In: *Communications of the ACM* 58 (2015), pp. 86–92. ISSN: 0001-0782, 1557-7317. DOI: 10.1145/2656334.

[308] Ladislao Salmerón, Yvonne Kammerer, and Pilar García-Carrión. "Searching the Web for conflicting topics: Page and user factors". In: *Computers in Human Behavior* 29.6 (2013), pp. 2161–2171.

[309] Hosam Al-Samarraie, Atef Eldenfria, and Husameddin Dawoud. "The Impact of Personality Traits on Users' Information-Seeking Behavior". In: *Information Processing & Management* 53 (2017), pp. 237–247. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2016.08.004.

[310] Mark Sanderson and W. Bruce Croft. "The History of Information Retrieval Research". In: *Proceedings of the IEEE* 100 (2012), pp. 1444–1451. ISSN: 1558-2256. DOI: 10.1109/JPROC.2012.2189916.

[311] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. *DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter*. 2019. DOI: 10.48550/ARXIV.1910.01108.

[312] Rodrygo L. T. Santos, Craig Macdonald, and Iadh Ounis. "Search Result Diversification". en. In: *Foundations and Trends® in Information Retrieval* 9.1 (2015), pp. 1–90. ISSN: 1554-0669, 1554-0677. DOI: 10.1561/1500000040. URL: http://www.nowpublishers.com/article/Details/INR-040 (visited on 07/13/2021).

[313] Piotr Sapiezynski, Wesley Zeng, Ronald E Robertson, Alan Mislove, and Christo Wilson. "Quantifying the Impact of User Attentionon Fair Group Representation in Ranked Lists". en. In: *Companion Proceedings of The 2019 World Wide Web Conference*. San Francisco USA: ACM, May 2019, pp. 553–562. ISBN: 978-1-4503-6675-5. DOI: 10.1145/3308560.3317595. (Visited on 07/13/2021).

[314] Reijo Savolainen. "Elaborating the Conceptual Space of Information-Seeking Phenomena". In: *Information Research: An International Electronic Journal* 21 (2016).

[315] Viola Schiaffonati et al. "Explorative Experiments and Digital Humanism: Adding an Epistemic Dimension to the Ethical Debate". In: *Perspectives on Digital Humanism* (2022), p. 77.

[316] Josephine B. Schmitt, Christina A. Debbelt, and Frank M. Schneider. "Too Much Information? Predictors of Information Overload in the Context of Online News Exposure". In: *Information, Communication & Society* 21 (2018), pp. 1151–1167. ISSN: 1369-118X, 1468-4462. DOI: 10.1080/1369118X.2017.1305427.

[317] Stefan Schweiger, Aileen Oeberst, and Ulrike Cress. "Confirmation Bias in Web-Based Search: A Randomized Online Study on the Effects of Expert Information and Social Tags on Information Search and Evaluation". In: *Journal of Medical Internet Research* 16 (2014), e3044. DOI: 10.2196/jmir.3044.

[318] Christina Schwind and Jürgen Buder. "Reducing Confirmation Bias and Evaluation Bias: When Are Preference-Inconsistent Recommendations Effective – and When Not?" In: *Computers in Human Behavior* 28 (2012), pp. 2280–2290. ISSN: 0747-5632. DOI: 10.1016/j.chb.2012.06.035.

[319] Oshani Seneviratne and James Hendler, eds. *Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web*. 1st ed. Vol. 52. New York, NY, USA: Association for Computing Machinery, 2023.

[320] Chirag Shah and Emily M. Bender. "Situating Search". In: *ACM SIGIR Conference on Human Information Interaction and Retrieval*. CHIIR '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 221–232. ISBN: 978-1-4503-9186-3. DOI: 10.1145/3498366.3505816.

[321] C. E. Shannon. "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[322] Boban Simonovic, Katia Vione, Dean Fido, Edward Stupple, James Martin, and Richard Clarke. "The Impact of Attitudes, Beliefs, and Cognitive Reflection on the Development of Critical Thinking Skills in Online Students". In: *Online Learning* 26 (2022). ISSN: 2472-5730. DOI: 10.24059/olj.v26i2.2725.

[323] Laura Slechten, Cédric Courtois, Lennert Coenen, and Bieke Zaman. "Adapting the Selective Exposure Perspective to Algorithmically Governed Platforms: The Case of Google Search". In: *Communication Research* 49 (2022), pp. 1039–1065. ISSN: 0093-6502, 1552-3810. DOI: 10.1177/00936502211012154.

[324] Annelien Smets, Lien Michiels, Toine Bogers, and Lennart Björneborn. "Serendipity in Recommender Systems Beyond the Algorithm: A Feature Repository and Experimental Design". In: *16th ACM Conference on Recommender Systems. CEUR Workshop Proceedings*. 2022, pp. 44–66.

[325] Catherine L. Smith and Soo Young Rieh. "Knowledge-Context in Search Systems: Toward Information-Literate Actions". In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. CHIIR '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 55–62. ISBN: 978-1-4503-6025-8. DOI: 10.1145/3295750.3298940.

[326] Jack B Soll, Katherine L Milkman, and John W Payne. "A User's Guide to Debiasing". In: *The Wiley Blackwell handbook of judgment and decision making* 2 (2015), pp. 924–951.

[327] Aaron Springer, Jean Garcia-Gathright, and Henriette Cramer. "Assessing and Addressing Algorithmic Bias-But Before We Get There..." In: *AAAI Spring Symposia*. 2018.

[328] Matthew L. Stanley, Alyssa H. Sinclair, and Paul Seli. "Intellectual Humility and Perceptions of Political Opponents". In: *Journal of Personality* 88 (2020), pp. 1196–1216. ISSN: 1467-6494. DOI: 10.1111/jopy.12566.

[329] Qingying Sun, Zhongqing Wang, Qiaoming Zhu, and Guodong Zhou. "Stance detection with hierarchical attention network". In: *Proceedings of the 27th international conference on computational linguistics*. 2018, pp. 2399–2409.

[330] Olof Sundin, Dirk Lewandowski, and Jutta Haider. "Whose Relevance? Web Search Engines as Multisided Relevance Machines". In: *Journal of the Association for Information Science and Technology* 73 (2022), pp. 637–642. ISSN: 2330-1643. DOI: 10.1002/asi.24570.

[331] Reid Swanson, Brian Ecker, and Marilyn Walker. "Argument Mining: Extracting Arguments from Online Dialogue". en. In: *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Prague, Czech Republic: Association for Computational Linguistics, 2015, pp. 217–226. DOI: 10.18653/v1/W15-4631. URL: http://aclweb.org/anthology/W15-4631 (visited on 07/13/2021).

[332] Bobby Swar, Tahir Hameed, and Iris Reychav. "Information overload, psychological ill-being, and behavioral intention to continue online healthcare information search". In: *Computers in human behavior* 70 (2017), pp. 416–425.

[333] Briony Swire-Thompson and David Lazer. "Public Health and Online Misinformation: Challenges and Recommendations". In: *Annual review of public health* 41 (2020), pp. 433–451. ISSN: 1545-2093. DOI: 10.1146/annurev-publhealth-040119-094127.

[334] Diana Tabatabai and Bruce M. Shore. "How Experts and Novices Search the Web". In: *Library & Information Science Research* 27 (2005), pp. 222–248. ISSN: 0740-8188. DOI: 10.1016/j.lisr.2005.01.005.

[335] Charles S. Taber, Damon Cann, and Simona Kucsova. "The Motivated Processing of Political Arguments". In: *Political Behavior* 31 (June 2009), pp. 137–155. ISSN: 1573-6687. DOI: 10.1007/s11109-008-9075-8.

[336] Herman Tavani. "Search Engines and Ethics". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2020. Metaphysics Research Lab, Stanford University, 2020. URL: https://plato.stanford.edu/archives/fall2020/entries/ethics-search/.

[337] Richard H Thaler and Cass R Sunstein. "Nudge: improving decisions about health". In: *Wealth, and Happiness* 6 (2008), pp. 14–38.

[338] Richard H Thaler and Cass R Sunstein. *Nudge: The final edition*. Yale University Press, 2021.

[339] Richard H. Thaler, Cass R. Sunstein, and John P. Balz. *Choice Architecture*. en. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, Apr. 2010. DOI: 10.2139/ssrn.1583509.

[340] Keela Thomson and Daniel Oppenheimer. "Investigating an Alternate Form of the Cognitive Reflection Test". In: *Judgment and Decision Making* 11 (2016), pp. 99–113. DOI: 10.1037/t49856-000.

[341] Thibaut Thonet, Guillaume Cabanac, Mohand Boughanem, and Karen Pinel-Sauvagnat. "Users Are Known by the Company They Keep: Topic Models for Viewpoint Discovery in Social Networks". en. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. Singapore Singapore: ACM, Nov. 2017, pp. 87–96. ISBN: 978-1-4503-4918-5. DOI: 10.1145/3132847.3132897. (Visited on 07/13/2021).

[342] Kjerstin Thorson and Chris Wells. "Curated Flows: A Framework for Mapping Media Exposure in the Digital Age". In: *Communication Theory* 26 (2016), pp. 309–328. ISSN: 1050-3293. DOI: 10.1111/comt.12087.

[343] Rob Tieben, Tilde Bekker, and Ben Schouten. "Curiosity and Interaction: Making People Curious through Interactive Systems". en. In: *Proceedings of HCI 2011 The 25th BCS Conference on Human Computer Interaction*. July 2011. DOI: 10.14236/ewic/HCI2011.66.

[344] Eran Toch, Yang Wang, and Lorrie Faith Cranor. "Personalization and Privacy: A Survey of Privacy Risks and Remedies in Personalization-Based Systems". In: *User Modeling and User-Adapted Interaction* 22 (2012), pp. 203–220. ISSN: 1573-1391. DOI: 10.1007/s11257-011-9110-z.

[345] Zakary L. Tormala and Derek D. Rucker. "Attitude Certainty: Antecedents, Consequences, and New Directions". In: *Consumer Psychology Review* 1 (2018), pp. 72–89. ISSN: 2476-1281. DOI: 10.1002/arcp.1004.

[346] Amine Trabelsi and Osmar R. Zaiane. "Finding Arguing Expressions of Divergent Viewpoints in Online Debates". en. In: *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)*. Gothenburg, Sweden: Association for Computational Linguistics, 2014, pp. 35–43. DOI: 10.3115/v1/W14-1305. URL: http://aclweb.org/anthology/W14-1305 (visited on 07/13/2021).

[347] Johanne R. Trippas, Damiano Spina, Paul Thomas, Mark Sanderson, Hideo Joho, and Lawrence Cavedon. "Towards a Model for Spoken Conversational Search". In: *Information Processing & Management* 57 (2020), p. 102162. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2019.102162.

[348] Yariv Tsfati and Joseph N. Cappella. "Why Do People Watch News They Do Not Trust? The Need for Cognition as a Moderator in the Association Between News Media Skepticism and Exposure". In: *Media Psychology* 7 (2005), pp. 251–271. ISSN: 1521-3269, 1532-785X. DOI: 10.1207/S1532785XMEP0703_2.

[349] Joshua A Tucker, Yannis Theocharis, Margaret E Roberts, and Pablo Barberá. "From liberation to turmoil: Social media and democracy". In: *J. Democracy* 28 (2017), p. 46.

[350] Amazon Mechanical Turk. 2021. URL: https://www.mturk.com/.

[351] Amos Tversky and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases". In: *Science* 185 (Sept. 1974), pp. 1124–1131. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.185.4157.1124.

[352] Aleksandra Urman, Mykola Makhortykh, and Roberto Ulloa. "The Matter of Chance: Auditing Web Search Results Related to the 2020 U.S. Presidential Primary Elections Across Six Search Engines". In: *Social Science Computer Review* 40 (2022), pp. 1323–1339. ISSN: 0894-4393. DOI: 10.1177/08944393211006863.

[353] Patti M Valkenburg and Jochen Peter. "The differential susceptibility to media effects model". In: *Journal of communication* 63.2 (2013), pp. 221–243.

[354] Elizabeth Van Couvering. "Is Relevance Relevant? Market, Science, and War: Discourses of Search Engine Quality". In: *Journal of Computer-Mediated Communication* 12 (2007), pp. 866–887. ISSN: 1083-6101. DOI: 10.1111/j.1083-6101.2007.00354.x.

[355] Lawrence Van den Bogaert, David Geerts, and Jaron Harambam. "Putting a Human Face on the Algorithm: Co-Designing Recommender Personae to Democratize News Recommender Systems". In: *Digital Journalism* (2022), pp. 1–21.

[356] Antal Van den Bosch, Toine Bogers, and Maurice De Kunder. "Estimating search engine index size variability: a 9-year longitudinal study". In: *Scientometrics* 107.2 (2016), pp. 839–856.

[357] Trevor Van Mierlo. "The 1% rule in four digital health social networks: an observational study". In: *Journal of medical Internet research* 16.2 (2014), e2966.

[358] Johan L. H. van Strien, Yvonne Kammerer, Saskia Brand-Gruwel, and Henny P. A. Boshuizen. "How Attitude Strength Biases Information Processing and Evaluation on the Web". In: *Computers in Human Behavior* 60 (2016), pp. 245–252. ISSN: 0747-5632. DOI: 10.1016/j.chb.2016.02.057.

[359] Liwen Vaughan and Mike Thelwall. "Search engine coverage bias: evidence and possible causes". In: *Information processing & management* 40.4 (2004), pp. 693–707.

[360] Dáša Vedejová and Vladimíra Čavojová. "Confirmation Bias in Information Search, Interpretation, and Memory Recall: Evidence from Reasoning about Four Controversial Topics". In: *Thinking & Reasoning* 28 (Jan. 2022), pp. 1–28. ISSN: 1354-6783. DOI: 10.1080/13546783.2021.1891967.

[361] Judith Vermeulen. "To Nudge or Not to Nudge: News Recommendation as a Tool to Achieve Online Media Pluralism". In: *Digital Journalism* (2022), pp. 1–20. ISSN: 2167-0811, 2167-082X. DOI: 10.1080/21670811.2022.2026796.

[362] Bas Verplanken, Pieter T Hazenberg, and Grace R Palenéwen. "Need for Cognition and External Information Search Effort". In: *Journal of Research in Personality* 26 (1992), pp. 128–136. ISSN: 0092-6566. DOI: 10.1016/0092-6566(92)90049-A.

[363] Sanne Vrijenhoek, Gabriel Bénédict, Mateo Gutierrez Granada, Daan Odijk, and Maarten De Rijke. "RADio–Rank-Aware Divergence Metrics to Measure Normative Diversity in News Recommendations". In: *Proceedings of the 16th ACM Conference on Recommender Systems*. 2022, pp. 208–219.

[364] Sanne Vrijenhoek, Mesut Kaya, Nadia Metoui, Judith Möller, Daan Odijk, and Natali Helberger. "Recommenders with a mission: assessing diversity in newsrecommendations". en. In: (Dec. 2020). arXiv: 2012.10185 [cs].

[365] Henning Wachsmuth, Martin Potthast, Khalid Al-Khatib, Yamen Ajjour, Jana Puschmann, Jiani Qu, Jonas Dorsch, Viorel Morari, Janek Bevendorff, and Benno Stein. "Building an Argument Search Engine for the Web". In: *Proceedings of the 4th Workshop on Argument Mining*. Copenhagen, Denmark: Association for Computational Linguistics, 2017, pp. 49–59. DOI: 10.18653/v1/W17-5106.

[366] Henning Wachsmuth, Shahbaz Syed, and Benno Stein. "Retrieval of the best counterargument without prior topic knowledge". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2018, pp. 241–251.

[367] Emily Wall, Leslie M Blaha, Lyndsey Franklin, and Alex Endert. "Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics". In: *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2017, pp. 104–115.

[368] Rui Wang, Deyu Zhou, Mingmin Jiang, Jiasheng Si, and Yang Yang. "A Survey on Opinion Mining: From Stance to Product Aspect". en. In: *IEEE Access* 7 (2019), pp. 41101–41124. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2906754. URL: https://ieeexplore.ieee.org/document/8672602/ (visited on 07/13/2021).

[369] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. "A Field Trial of Privacy Nudges for Facebook". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. New York, NY, USA: Association for Computing Machinery, 2014, pp. 2367–2376. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557413.

[370] Ingmar Weber and Alejandro Jaimes. "Who Uses Web Search for What: And How". In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. WSDM '11. New York, NY, USA: Association for Computing Machinery, 2011, pp. 15–24. ISBN: 978-1-4503-0493-1. DOI: 10.1145/1935826.1935839.

[371] Hannes Werthner, Erich Prem, Edward A. Lee, and Carlo Ghezzi, eds. *Perspectives on Digital Humanism*. Springer Nature, 2022. DOI: 10.1007/978-3-030-86144-5.

[372] Ryen White. "Beliefs and Biases in Web Search". In: *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '13. New York, NY, USA: Association for Computing Machinery, 2013, pp. 3–12. ISBN: 978-1-4503-2034-4. DOI: 10.1145/2484028.2484053.

[373] Ryen W. White. "Belief Dynamics in Web Search". In: *Journal of the Association for Information Science and Technology* 65 (2014), pp. 2165–2178. ISSN: 2330-1643. DOI: 10.1002/asi.23128.

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

[375] Ryen W. White and Ahmed Hassan. "Content Bias in Online Health Search". In: *ACM Transactions on the Web* 8 (2014), 25:1–25:33. ISSN: 1559-1131. DOI: 10.1145/2663355.

[376] Ryen W. White and Eric Horvitz. "Belief Dynamics and Biases in Web Search". In: *ACM Transactions on Information Systems* 33 (2015), 18:1–18:46. ISSN: 1046-8188. DOI: 10.1145/2746229.

[377] Ryen W. White and Resa A. Roth. "Exploratory Search: Beyond the Query-Response Paradigm". In: *Synthesis Lectures on Information Concepts, Retrieval, and Services* 1 (2009), pp. 1–98. ISSN: 1947-945X. DOI: 10.2200/S00174ED1V01Y200901ICR003.

[378] Wikipedia. *Confirmation bias*. May 2021. URL: https://en.wikipedia.org/wiki/Confirmation_bias.

[379] Ben James Winer, Donald R Brown, and Kenneth M Michels. *Statistical principles in experimental design*. Vol. 2. McGraw-Hill New York, 1971.

[380] Susan J Winter and Brian S Butler. "Responsible technology design: Conversations for success". In: *Perspectives on Digital Humanism* (2022), pp. 271–275.

[381] J David Wolfgang, Tim P Vos, Kimberly Kelling, and Sooyoung Shin. "Political Journalism and Democracy: How Journalists Reflect Political Viewpoint Diversity in Their Reporting". In: *Journalism Studies* 22.10 (2021), pp. 1339–1357.

[382] Zhangyi Wu, Tim Draws, Federico Cau, Francesco Barile, **Alisa Rieger**, and Nava Tintarev. "Explaining Search Result Stances to Opinionated People". In: *Explainable Artificial Intelligence*. Ed. by Luca Longo. Cham: Springer Nature Switzerland, 2023, pp. 573–596. ISBN: 978-3-031-44067-0. DOI: 10.1007/978-3-031-44067-0_29.

[383] Luyan Xu, Mengdie Zhuang, and Ujwal Gadiraju. "How Do User Opinions Influence Their Interaction With Web Search Results?" In: *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 240–244. ISBN: 978-1-4503-8366-0. DOI: 10.1145/3450613.3456824.

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

[385] Ke Yang and Julia Stoyanovich. "Measuring Fairness in Ranked Outputs". en. In: *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*. Chicago IL USA: ACM, June 2017, pp. 1–6. ISBN: 978-1-4503-5282-6. DOI: 10.1145/3085504.3085526. (Visited on 07/13/2021).

[386] Ke Yang, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, Hv Jagadish, and Gerome Miklau. "A Nutritional Label for Rankings". en. In: *Proceedings of the 2018 International Conference on Management of Data*. Houston TX USA: ACM, May 2018, pp. 1773–1776. ISBN: 978-1-4503-4703-7. DOI: 10.1145/3183713.3193568. (Visited on 07/13/2021).

[387] Dawei Yin, Yuening Hu, Jiliang Tang, Tim Daly, Mianwei Zhou, Hua Ouyang, Jianhui Chen, Changsung Kang, Hongbo Deng, Chikashi Nobata, Jean-Marc Langlois, and Yi Chang. "Ranking Relevance in Yahoo Search". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 323–332. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939677.

[388] Elad Yom-Tov, Susan Dumais, and Qi Guo. "Promoting Civil Discourse Through Search Engine Diversity". en. In: *Social Science Computer Review* 32.2 (Apr. 2014), pp. 145–154. ISSN: 0894-4393, 1552-8286. DOI: 10.1177/0894439313506838. (Visited on 07/13/2021).

[389] Ran Yu, Ujwal Gadiraju, Peter Holtz, Markus Rokicki, Philipp Kemkes, and Stefan Dietze. "Predicting user knowledge gain in informational search sessions". In: *The 41st international ACM SIGIR conference on research & development in information retrieval*. 2018, pp. 75–84.

[390] Yisong Yue, Rajan Patel, and Hein Roehrig. "Beyond position bias: Examining result attractiveness as a source of presentation bias in clickthrough data". In: *Proceedings of the 19th international conference on World wide web*. 2010, pp. 1011–1018.

[391] Corinne E Zachry, Le Vy Phan, Laura E R Blackie, and Eranda Jayawickreme. "Situation-Based Contingencies Underlying Wisdom-Content Manifestations: Examining Intellectual Humility in Daily Life". In: *The Journals of Gerontology: Series B* 73 (2018), pp. 1404–1415. ISSN: 1079-5014. DOI: 10.1093/geronb/gby016.

[392] Hamed Zamani, Michael Bendersky, Xuanhui Wang, and Mingyang Zhang. "Situational Context for Ranking in Personal Search". In: *Proceedings of the 26th International Conference on World Wide Web*. WWW '17. Perth, Australia: International World Wide Web Conferences Steering Committee, 2017, pp. 1531–1540. ISBN: 9781450349130. DOI: 10.1145/3038912.3052648.

[393] Meike Zehlike, Ke Yang, and Julia Stoyanovich. "Fairness in ranking, part I: Score-based ranking". In: *ACM Computing Surveys* 55.6 (2022), pp. 1–36.

[394] Dake Zhang, Amir Vakili Tahami, Mustafa Abualsaud, and Mark D. Smucker. "Learning Trustworthy Web Sources to Derive Correct Answers and Reduce Health Misinformation in Search". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 2099–2104. ISBN: 978-1-4503-8732-3. DOI: 10.1145/3477495.3531812.

[395] Kaiping Zheng, Hongzhi Wang, Zhixin Qi, Jianzhong Li, and Hong Gao. "A survey of query result diversification". In: *Knowledge and Information Systems* 51.1 (2017), pp. 1–36.

[396] Xinyi Zhou and Reza Zafarani. "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities". In: *ACM Computing Surveys* 53 (2020), 109:1–109:40. ISSN: 0360-0300. DOI: 10.1145/3395046.

[397] Lixing Zhu, Yulan He, and Deyu Zhou. "Hierarchical viewpoint discovery from tweets using Bayesian modelling". en. In: *Expert Systems with Applications* 116 (Feb. 2019), pp. 430–438. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2018.09.028. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957417418306055 (visited on 07/13/2021).

[398] Steven Zimmerman, Stefan M. Herzog, David Elsweiler, Jon Chamberlain, and Udo Kruschwitz. "Towards a Framework for Harm Prevention in Web Search". In: *Proceedings of the First Workshop on Bridging the Gap between Information Science, Information Retrieval and Data Science (BIRDS 2020), Co-Located with 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 202*. Ed. by Ingo Frommholz, Haiming Liu, and Massimo Melucci. Vol. 2741. Xi'an, China (online).: CEUR Workshop Proceedings, 2020, pp. 30–46.

[399] Leor Zmigrod, Sharon Zmigrod, Peter Jason Rentfrow, and Trevor W. Robbins. "The Psychological Roots of Intellectual Humility: The Role of Intelligence and Cognitive Flexibility". In: *Personality and Individual Differences* 141 (2019), pp. 200–208. ISSN: 0191-8869. DOI: 10.1016/j.paid.2019.01.016.

SUMMARIES

ENGLISH SUMMARY

Web search plays an important role in the contemporary information landscape, shaping individual and collective knowledge by providing fast and effortless access to vast amounts of resources. We rely on web search engines for various information needs, some of which can carry serious consequences. This is particularly evident when searching for information on debated topics, which can shape opinions and practical decisions. Debated topics are characterized by diverse and often opposing perspectives linked to different values and interests. Ideally, individuals would diligently engage with different perspectives to become well-informed and form opinions responsibly. However, engaging with information on debated topics can be cognitively demanding and subject to emotionally charged and biased behavior. When resorting to web search to find information on debated topics, searchers may be confronted with further obstacles. For instance, search engines are known to apply opaque ranking criteria, may not provide sufficient viewpoint diversity, and might foster over-reliance.

In this dissertation, we present different user studies aimed at better understanding the challenges of web search on debated topics and identifying measures to help searchers overcome these challenges. We first explored whether and how factors inherent to the searcher and search interface affect search behavior. Then, we investigated the risks and benefits of interventions to guide search behavior as well as empower searchers, aiming at supporting unbiased and diligent search interactions without restricting searcher autonomy. Our findings underscore the unique characteristics of web search on debated topics and provide a foundation for designing, tailoring, and evaluating interventions to support searchers. Considering the overall insights gained through our user studies, it becomes clear that the most pivotal challenges of web search on debated topics arise from the complex searcher-system interplay. Rather than turning to simple fixes, there is a need to acknowledge the complexity of the issue and commit to comprehensive investigations and solutions to avoid inadvertently exacerbating risks. Laying the groundwork for future investigations, we provide an extensive review of interdisciplinary literature with a detailed account of challenges and research opportunities.

With this dissertation, we raise awareness for the pressing socio-technical issues related to digital media and opinion formation and aspire to encourage interdisciplinary research teams, practitioners, and policymakers to join forces in establishing web search environments that foster individual and societal well-being.

NEDERLANDSE SAMENVATTING

Zoeken op het web speelt een belangrijke rol bij het verkrijgen van informatie. Door snelle en moeiteloze toegang te bieden tot enorme hoeveelheden bronnen, geeft het vorm aan individuele en collectieve kennis. We gebruiken zoekmachines voor verschillende taken, waaronder taken die verder gaan dan het eenvoudig ophalen van informatie en die belangrijke gevolgen kunnen hebben. Dit is vooral duidelijk bij het zoeken naar informatie over omstreden onderwerpen die meningen en praktische beslissingen kunnen beïnvloeden. Omstreden onderwerpen worden gekenmerkt door verschillende, vaak tegenstrijdige perspectieven die gekoppeld zijn aan verschillende waarden en belangen. Idealiter zouden zoekers de verschillende perspectieven zorgvuldig overwegen om goed geïnformeerd te zijn en een verantwoorde mening te kunnen vormen. Zich bezighouden met dergelijke onderwerpen kan echter cognitief veeleisend zijn en leiden tot emotioneel en bevooroordeeld gedrag. Als zoekmachines voor dit doel worden gebruikt, gaat dit vaak gepaard met nog meer obstakels. Ze passen bijvoorbeeld meestal ondoorzichtige rangschikkingscriteria toe, weerspiegelen de diversiteit van meningen niet adequaat en worden ervan verdacht een overmatig vertrouwen te veroorzaken.

In deze dissertatie presenteren we verschillende gebruikersonderzoeken om de uitdagingen van het zoeken op het web naar omstreden onderwerpen beter te begrijpen en om maatregelen te identificeren die zoekers kunnen helpen deze uitdagingen te overwinnen. Eerst onderzochten we of en hoe kenmerken van zoekers en de zoekinterface zoekgedrag beïnvloeden. Vervolgens onderzochten we de voor- en nadelen van interventies om zoekgedrag te sturen en zoekers te empoweren. Onze bevindingen onthullen unieke kenmerken van het zoeken op het web naar omstreden onderwerpen en bieden een basis voor het ontwikkelen, aanpassen en onderzoeken van interventies om zoekers te ondersteunen. In het licht van al onze bevindingen is het duidelijk dat de grootste uitdagingen voortkomen uit de complexe wisselwerking tussen zoekers en zoekmachines. In plaats van te kiezen voor eenvoudige oplossingen die onbedoeld de risico's kunnen vergroten, moet de complexiteit worden erkend. Als basis voor essentieel vervolgonderzoek presenteren we de uitdagingen en mogelijkheden die we hebben geïdentificeerd door middel van interdisciplinair literatuuronderzoek.

Met dit proefschrift vergroten we het bewustzijn van de urgente socio-technische kwesties met betrekking tot digitale media en opinievorming. We willen interdisciplinaire onderzoeksteams en besluitvormers in het bedrijfsleven, de juridische wereld en de politiek aanmoedigen om samen te werken aan het creëren van webzoekomgevingen die het individuele en maatschappelijke welzijn bevorderen.

DEUTSCHE ZUSAMMENFASSUNG

Websuche spielt eine wichtige Rolle in der Informationsbeschaffung. Durch schnellen und mühelosen Zugang zu enormen Mengen an Ressourcen, prägen Suchmaschinen individuelles und kollektives Wissen. Wir nutzen Suchmaschinen für diverse Aufgaben, auch solche die über das einfache Abrufen von Information hinausgehen und erhebliche Konsequenzen haben können. Das wird besonders deutlich bei Suchen nach Informationen zu debattierten Themen, die Meinungen und praktische Entscheidungen beeinflussen können. Debattierte Themen sind geprägt von verschiedenen, oft gegensätzlichen Perspektiven, die mit unterschiedlichen Werten und Interessen verbunden sind. Idealerweise würden sich Suchende sorgfältig mit verschiedenen Perspektiven auseinandersetzen, um für eine verantwortungsvolle Meinungsbildung gut informiert zu sein. Sich mit solchen Themen auseinanderzusetzen kann allerdings kognitiv anspruchsvoll sein und zu emotionalem und voreingenommenem Verhalten führen. Wenn dafür Suchmaschinen genutzt werden, geht das oft mit weiteren Hindernissen einher. Zum Beispiel wenden sie meistens undurchsichtige Rankingkriterien an, bilden Meinungsvielfalt nicht ausreichend ab und stehen im Verdacht, übermäßiges Vertrauen hervorzurufen.

In dieser Dissertation präsentieren wir verschiedene Nutzerstudien, mit denen wir die Herausforderungen der Websuche zu debattierten Themen besser verstehen und Maßnahmen finden wollten, um Suchende bei der Bewältigung dieser Herausforderungen zu unterstützen. Zunächst haben wir untersucht, ob und wie Charakteristika der Suchenden und des Suchinterfaces das Suchverhalten beeinflussen. Im Anschluss haben wir die Vor- und Nachteile von Maßnahmen zur Steuerung des Suchverhaltens und zur Stärkung der Fähigkeiten der Suchenden untersucht. Unsere Ergebnisse zeigen einzigartige Eigenschaften von Websuche zu debattierten Themen und bieten eine Grundlage für die Entwicklung, Anpassung und Erforschung von Maßnahmen zur Unterstützung von Suchenden. In Anbetracht aller unserer Erkenntnisse wird deutlich, dass die größten Herausforderungen aus dem komplexen Zusammenspiel von Suchenden und Suchmaschinen resultieren. Anstatt sich einfachen Lösungen zuzuwenden, die Risiken unabsichtlich vergrößern könnten, muss die Komplexität anerkannt werden. Als Grundlage für zukünftige Untersuchungen haben wir die Herausforderungen und Forschungsmöglichkeiten mit einer interdisziplinären Literaturrecherche herausgearbeitet.

Mit dieser Dissertation schärfen wir das Bewusstsein für die drängenden sozio-technischen Probleme im Zusammenhang mit digitalen Medien und Meinungsbildung. Wir wollen interdisziplinäre Forschungsteams und Entscheidungstragende in Wirtschaft, Recht und Politik dazu ermutigen, sich gemeinsam für die Schaffung von Websuchumgebungen einzusetzen, die das individuelle und gesellschaftliche Wohlbefinden fördern.

ACKNOWLEDGEMENTS

This dissertation would not exist without the direct and indirect support of many wonderful people, to whom I owe a huge thank you.

Thank you to Dr. Jahna Otterbacher, Prof. Djoerd Hiemstra, Prof. Ibo van de Poel, Prof. Catholijn Jonker, and Dr. Avishek Anand, for making time to be part of my doctoral committee, reviewing my dissertation, and for your insightful feedback.

I would like to express my appreciation to everyone who was directly involved in my supervision. A huge thank you to my co-promotor Sole. Thank you for guiding me steadily and with affection through the intense final PhD years, always having your door open, encouraging me to do my best, giving me tons and tons of feedback, leaving little treats with motivational notes on my desk, and sharing moments of enthusiasm and excitement, as well as disappointment and frustration. Despite the pressure of the tight timeline, it has been a lot of fun. You are an exceptional supervisor and I consider myself very lucky that you joined WIS and were there to guide me through the last years. I want to express my gratitude to my promotor, Geert-Jan, for many insightful discussions and for providing advice and support throughout the smooth and rough patches of my PhD journey. A heartfelt thank you to Mariët for patiently listening to my ideas, plans, or concerns and sharing reflections and advice. Your unwavering support and remarkable kindness have been deeply appreciated. I also would like to acknowledge Nava, who gave me the opportunity to do this PhD and guided me through a big part of this journey, particularly the challenging initial years overshadowed by the COVID-19 pandemic. Each of you has played an important role in my academic and personal growth.

Thank you to everyone in WIS, especially the Epsilon/XAI and Kappa teams, for the discussions, continuous feedback, and collaborations throughout the past four years. A special thank you to Tim. Your expertise in conducting and evaluating user studies, coupled with your patient guidance, has greatly contributed to the refinement of my research skills and our many conversations on search on debated topics, responsible opinions, and biases are reflected in the very foundations of this dissertation. I am deeply grateful to Ujwal, who challenged and encouraged me in equal measures during countless bike ride discussions. Co-supervising master's theses with you, I have learned a lot, from conducting HCI research to kindling students' enthusiasm and guiding them with genuine kindness. It has been a great pleasure to work with you, please greet the Heron from me on your bike rides to come! Thank you to Daphne and Nadia for the invaluable support with the many organizational tasks throughout this PhD, from making me feel welcome and assisting me in the transition before my contract started to organizing my defense.

In addition to my supervisory team and wonderful WIS colleagues, this dissertation is a testimony to collaborative efforts with other talented individuals who I am honored to call my co-authors. I especially want to thank Suleiman, Frank, Zhangyi, Nick, David, Federico, and Francesco for their contributions to shaping some of the research projects detailed in this dissertation.

Having had the privilege of doing my PhD as part of a Marie Skłodowska Curie Network, I got to work with and receive training, guidance, and advice from a multitude of excellent researchers from different institutions across Europe. I am grateful to everyone involved in the project for all the exchange, input, support, and fun. Special thanks to Raluca, for being absolutely amazing, as well as the HMI group at the University of Twente, the IIIA-CSIC, and Wizenoze, particularly Mariët and Sumit, Carles and Qurat, and Thijs for hosting me during my secondments.

Reflecting on the past four years, I am most grateful for all the wonderful people I got to meet in Delft and specifically WIS, the NL4XAI project, during secondments, and at conferences. The many memorable moments we shared and the friendships that have grown from them have been the greatest rewards of this journey. Online coffee meetings, crappy Echo lunches, library coffee walks, Doerak vegan nuggets, evenings at Koornmarkt, NDT performances, queuing for the very best ramen, Den Haag beach trips, kayaking and carting, Amsterdam barbecues, the IIIA Costellada, HMI and Lumen movie nights, Tankstation memory matches, summer evenings with amazing risotto in Utrecht, waffle breakfasts in Warsaw, pool tournaments in Nancy and Enschede, culinary treats in Santiago de Compostela, searching and never finding karaoke bars, late-night strolls through Barcelona, King's day manicure in Hamburg, Highland hiking adventures, flamingo spotting in Cagliari, Barbie expeditions to Rome and Romania, Lirac vineyard walks, and Elli-sitting in Dannenrod. All that and more in your exquisite company contributed substantially to my happiness and, consequentially, the completion of this dissertation. You exceed by far what I could have wished for. A special thank you to my sweet paranymphs, Andra and Christos.

Finally, I want to thank my magnificent friends and family. Kris, thank you for accompanying, supporting, and inspiring me along my journey and adding a lot of happiness to it. Also, thanks for encouraging me, on the day of the application deadline and while we were on vacation, to apply for this PhD position that I had already dismissed. Mara, thank you for designing the beautiful cover of this dissertation and being there for every chapter of my life over the past 25 years, and Franzi, thank you for convincing me that doing research indeed can be enjoyable and rewarding. To my parents, Mama und Papa, and brothers, Janek, Luis, and Elias, thank you for being my secure base and equipping me with curiosity, courage, a critical mind, appreciation of the good things, and lots of love.

Curriculum Vitæ

Alisa Rieger

06-02-1995	Born in Erbach, Germany

Education

2020–2024	Doctor of Philosopy (PhD), Computer Science Delft University of Technology, Netherlands
2017–2020	Master of Science (MSc), Sensors and Cognitive Psychology Chemnitz University of Technology, Germany
2014-2017	Bachelor of Science (BSc), Sensors and Cognitive Psychology Chemnitz University of Technology, Germany

Awards

2021	Douglas Engelbart Best Paper Award at ACM Hypertext Conference
2021	Amazon Best Paper Award at AAAI HCOMP Conference
2021	Delft Design for Value Open Subsidy

List of Publications

1. **Alisa Rieger**, Suleiman Kulane, Ujwal Gadiraju, and Maria Soledad Pera. "Disentangling Web Search on Debated Topics: A User-Centered Exploration". In: *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 24–35. ISBN: 9798400704338. DOI: 10.1145/3627043.3659559

2. **Alisa Rieger**, Frank Bredius, Mariët Theune, and Maria Soledad Pera. "From Potential to Practice: Intellectual Humility During Search on Debated Topics". In: *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*. CHIIR '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 130–141. DOI: 10.1145/3627508.3638306

3. **Alisa Rieger**, Tim Draws, Nicolas Mattis, David Maxwell, David Elsweiler, Ujwal Gadiraju, Dana McKay, Alessandro Bozzon, and Maria Soledad Pera. "Responsible Opinion Formation on Debated Topics in Web Search". In: *Advances in Information Retrieval*. Ed. by Nazli Goharian, Nicola Tonellotto, Yulan He, Aldo Lipani, Graham McDonald, Craig Macdonald, and Iadh Ounis. Cham: Springer Nature Switzerland, 2024, pp. 437–465. ISBN: 978-3-031-56066-8. DOI: 10.1007/978-3-031-56066-8_32

4. **Alisa Rieger**, Tim Draws, Mariët Theune, and Nava Tintarev. "Nudges to Mitigate Confirmation Bias during Web Search on Debated Topics: Support vs. Manipulation". In: *ACM Transactions on the Web* (2023). DOI: 10.1145/3635034

5. Zhangyi Wu, Tim Draws, Federico Cau, Francesco Barile, **Alisa Rieger**, and Nava Tintarev. "Explaining Search Result Stances to Opinionated People". In: *Explainable Artificial Intelligence*. Ed. by Luca Longo. Cham: Springer Nature Switzerland, 2023, pp. 573–596. ISBN: 978-3-031-44067-0. DOI: 10.1007/978-3-031-44067-0_29

6. **Alisa Rieger**, Frank Bredius, Nava Tintarev, and Maria Soledad Pera. "Searching for the Whole Truth: Harnessing the Power of Intellectual Humility to Boost Better Search on Debated Topics". In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI EA '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 1–8. ISBN: 978-1-4503-9422-2. DOI: 10.1145/3544549.3585693

7. Tim Draws, Nirmal Roy, Oana Inel, **Alisa Rieger**, Rishav Hada, Mehmet Orcun Yalcin, Benjamin Timmermans, and Nava Tintarev. "Viewpoint Diversity in Search Results". In: *Advances in Information Retrieval*. Ed. by Jaap Kamps, Lorraine Goeuriot, Fabio Crestani, Maria Maistro, Hideo Joho, Brian Davis, Cathal Gurrin, Udo Kruschwitz, and Annalina Caputo. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023, pp. 279–297. ISBN: 978-3-031-28244-7. DOI: 10.1007/978-3-031-28244-7_18

8. Francesco Barile, Tim Draws, Oana Inel, **Alisa Rieger**, Shabnam Najafian, Amir Ebrahimi Fard, Rishav Hada, and Nava Tintarev. "Evaluating Explainable Social

Choice-Based Aggregation Strategies for Group Recommendation". In: *User Modeling and User-Adapted Interaction* 34 (2024), pp. 1–58. ISSN: 1573-1391. DOI: 10.1007/s11257-023-09363-0

9. Alejandra Bringas Colmenarejo, Luca Nannini, **Alisa Rieger**, Kristen M. Scott, Xuan Zhao, Gourab K Patro, Gjergji Kasneci, and Katharina Kinder-Kurlanda. "Fairness in Agreement With European Values: An Interdisciplinary Perspective on AI Regulation". In: *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. AIES '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 107–118. ISBN: 978-1-4503-9247-1. DOI: 10.1145/3514094.3534158

10. **Alisa Rieger**. "Interactive Interventions to Mitigate Cognitive Bias". In: *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 316–320. ISBN: 978-1-4503-9207-5. DOI: 10.1145/3503252.3534362

11. **Alisa Rieger**, Qurat-Ul-Ain Shaheen, Carles Sierra, Mariet Theune, and Nava Tintarev. "Towards Healthy Engagement with Online Debates: An Investigation of Debate Summaries and Personalized Persuasive Suggestions". In: *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '22 Adjunct. New York, NY, USA: Association for Computing Machinery, 2022, pp. 192–199. ISBN: 978-1-4503-9232-7. DOI: 10.1145/3511047.3537692

12. **Alisa Rieger**, Tim Draws, Mariët Theune, and Nava Tintarev. "This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias". In: *Proceedings of the 32st ACM Conference on Hypertext and Social Media*. Virtual Event USA: ACM, Aug. 2021, pp. 189–199. ISBN: 978-1-4503-8551-0. DOI: 10.1145/3465336.3475101

13. Francesco Barile, Shabnam Najafian, Tim Draws, Oana Inel, **Alisa Rieger**, Rishav Hada, and Nava Tintarev. "Toward Benchmarking Group Explanations: Evaluating the Effect of Aggregation Strategies versus Explanation." In: *Perspectives@ RecSys*. 2021

14. Tim Draws, **Alisa Rieger**, Oana Inel, Ujwal Gadiraju, and Nava Tintarev. "A Checklist to Combat Cognitive Biases in Crowdsourcing". In: *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 9 (2021), pp. 48–59. ISSN: 2769-1349. DOI: 10.1609/hcomp.v9i1.18939

15. **Alisa Rieger**, Mariët Theune, and Nava Tintarev. "Toward Natural Language Mitigation Strategies for Cognitive Biases in Recommender Systems". In: *2nd Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence*. Dublin, Ireland: Association for Computational Linguistics, Nov. 2020, pp. 50–54

16. Laurianne Charrier, **Alisa Rieger**, Alexandre Galdeano, Amélie Cordier, Mathieu Lefort, and Salima Hassas. "The RoPE Scale: A Measure of How Empathic a Robot Is Perceived". In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2019, pp. 656–657. DOI: 10.1109/HRI.2019.8673082

SIKS Dissertation Series

Since 1998, all dissertations written by PhD. students who have conducted their research under auspices of a senior research fellow of the SIKS research school are published in the SIKS Dissertation Series (following are all the dissertations since 2016).

2016 01 Syed Saiden Abbas (RUN), Recognition of Shapes by Humans and Machines
02 Michiel Christiaan Meulendijk (UU), Optimizing medication reviews through decision support: prescribing a better pill to swallow
03 Maya Sappelli (RUN), Knowledge Work in Context: User Centered Knowledge Worker Support
04 Laurens Rietveld (VU), Publishing and Consuming Linked Data
05 Evgeny Sherkhonov (UVA), Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers
06 Michel Wilson (TUD), Robust scheduling in an uncertain environment
07 Jeroen de Man (VU), Measuring and modeling negative emotions for virtual training
08 Matje van de Camp (TiU), A Link to the Past: Constructing Historical Social Networks from Unstructured Data
09 Archana Nottamkandath (VU), Trusting Crowdsourced Information on Cultural Artefacts
10 George Karafotias (VUA), Parameter Control for Evolutionary Algorithms
11 Anne Schuth (UVA), Search Engines that Learn from Their Users
12 Max Knobbout (UU), Logics for Modelling and Verifying Normative Multi-Agent Systems
13 Nana Baah Gyan (VU), The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach
14 Ravi Khadka (UU), Revisiting Legacy Software System Modernization
15 Steffen Michels (RUN), Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments
16 Guangliang Li (UVA), Socially Intelligent Autonomous Agents that Learn from Human Reward
17 Berend Weel (VU), Towards Embodied Evolution of Robot Organisms
18 Albert Meroño Peñuela (VU), Refining Statistical Data on the Web
19 Julia Efremova (Tu/e), Mining Social Structures from Genealogical Data
20 Daan Odijk (UVA), Context & Semantics in News & Web Search
21 Alejandro Moreno Célleri (UT), From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground
22 Grace Lewis (VU), Software Architecture Strategies for Cyber-Foraging Systems
23 Fei Cai (UVA), Query Auto Completion in Information Retrieval
24 Brend Wanders (UT), Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach
25 Julia Kiseleva (TU/e), Using Contextual Information to Understand Searching and Browsing Behavior
26 Dilhan Thilakarathne (VU), In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains
27 Wen Li (TUD), Understanding Geo-spatial Information on Social Media
28 Mingxin Zhang (TUD), Large-scale Agent-based Social Simulation - A study on epidemic prediction and control
29 Nicolas Höning (TUD), Peak reduction in decentralised electricity systems - Markets and prices for flexible planning
30 Ruud Mattheij (UvT), The Eyes Have It
31 Mohammad Khelghati (UT), Deep web content monitoring
32 Eelco Vriezekolk (UT), Assessing Telecommunication Service Availability Risks for Crisis Organisations
33 Peter Bloem (UVA), Single Sample Statistics, exercises in learning from just one example
34 Dennis Schunselaar (TUE), Configurable Process Trees: Elicitation, Analysis, and Enactment
35 Zhaochun Ren (UVA), Monitoring Social Media: Summarization, Classification and Recommendation
36 Daphne Karreman (UT), Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies
37 Giovanni Sileno (UvA), Aligning Law and Action - a conceptual and computational inquiry
38 Andrea Minuto (UT), Materials that Matter - Smart Materials meet Art & Interaction Design
39 Merijn Bruijnes (UT), Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect

40 Christian Detweiler (TUD), Accounting for Values in Design
41 Thomas King (TUD), Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance
42 Spyros Martzoukos (UVA), Combinatorial and Compositional Aspects of Bilingual Aligned Corpora
43 Saskia Koldijk (RUN), Context-Aware Support for Stress Self-Management: From Theory to Practice
44 Thibault Sellam (UVA), Automatic Assistants for Database Exploration
45 Bram van de Laar (UT), Experiencing Brain-Computer Interface Control
46 Jorge Gallego Perez (UT), Robots to Make you Happy
47 Christina Weber (UL), Real-time foresight - Preparedness for dynamic innovation networks
48 Tanja Buttler (TUD), Collecting Lessons Learned
49 Gleb Polevoy (TUD), Participation and Interaction in Projects. A Game-Theoretic Analysis
50 Yan Wang (UVT), The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

2017 01 Jan-Jaap Oerlemans (UL), Investigating Cybercrime
02 Sjoerd Timmer (UU), Designing and Understanding Forensic Bayesian Networks using Argumentation
03 Daniël Harold Telgen (UU), Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines
04 Mrunal Gawade (CWI), Multi-core Parallelism in a Column-store
05 Mahdieh Shadi (UVA), Collaboration Behavior
06 Damir Vandic (EUR), Intelligent Information Systems for Web Product Search
07 Roel Bertens (UU), Insight in Information: from Abstract to Anomaly
08 Rob Konijn (VU) , Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery
09 Dong Nguyen (UT), Text as Social and Cultural Data: A Computational Perspective on Variation in Text
10 Robby van Delden (UT), (Steering) Interactive Play Behavior
11 Florian Kunneman (RUN), Modelling patterns of time and emotion in Twitter #anticipointment
12 Sander Leemans (TUE), Robust Process Mining with Guarantees
13 Gijs Huisman (UT), Social Touch Technology - Extending the reach of social touch through haptic technology
14 Shoshannah Tekofsky (UvT), You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior
15 Peter Berck (RUN), Memory-Based Text Correction
16 Aleksandr Chuklin (UVA), Understanding and Modeling Users of Modern Search Engines
17 Daniel Dimov (UL), Crowdsourced Online Dispute Resolution
18 Ridho Reinanda (UVA), Entity Associations for Search
19 Jeroen Vuurens (UT), Proximity of Terms, Texts and Semantic Vectors in Information Retrieval
20 Mohammadbashir Sedighi (TUD), Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility
21 Jeroen Linssen (UT), Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)
22 Sara Magliacane (VU), Logics for causal inference under uncertainty
23 David Graus (UVA), Entities of Interest — Discovery in Digital Traces
24 Chang Wang (TUD), Use of Affordances for Efficient Robot Learning
25 Veruska Zamborlini (VU), Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search
26 Merel Jung (UT), Socially intelligent robots that understand and respond to human touch
27 Michiel Joosse (UT), Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors
28 John Klein (VU), Architecture Practices for Complex Contexts
29 Adel Alhuraibi (UvT), From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT"
30 Wilma Latuny (UvT), The Power of Facial Expressions
31 Ben Ruijl (UL), Advances in computational methods for QFT calculations
32 Thaer Samar (RUN), Access to and Retrievability of Content in Web Archives
33 Brigit van Loggem (OU), Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity
34 Maren Scheffel (OU), The Evaluation Framework for Learning Analytics
35 Martine de Vos (VU), Interpreting natural science spreadsheets
36 Yuanhao Guo (UL), Shape Analysis for Phenotype Characterisation from High-throughput Imaging
37 Alejandro Montes Garcia (TUE), WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy
38 Alex Kayal (TUD), Normative Social Applications
39 Sara Ahmadi (RUN), Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR

40 Altaf Hussain Abro (VUA), Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems
41 Adnan Manzoor (VUA), Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle
42 Elena Sokolova (RUN), Causal discovery from mixed and missing data with applications on ADHD datasets
43 Maaike de Boer (RUN), Semantic Mapping in Video Retrieval
44 Garm Lucassen (UU), Understanding User Stories - Computational Linguistics in Agile Requirements Engineering
45 Bas Testerink (UU), Decentralized Runtime Norm Enforcement
46 Jan Schneider (OU), Sensor-based Learning Support
47 Jie Yang (TUD), Crowd Knowledge Creation Acceleration
48 Angel Suarez (OU), Collaborative inquiry-based learning

2018 01 Han van der Aa (VUA), Comparing and Aligning Process Representations
02 Felix Mannhardt (TUE), Multi-perspective Process Mining
03 Steven Bosems (UT), Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction
04 Jordan Janeiro (TUD), Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks
05 Hugo Huurdeman (UVA), Supporting the Complex Dynamics of the Information Seeking Process
06 Dan Ionita (UT), Model-Driven Information Security Risk Assessment of Socio-Technical Systems
07 Jieting Luo (UU), A formal account of opportunism in multi-agent systems
08 Rick Smetsers (RUN), Advances in Model Learning for Software Systems
09 Xu Xie (TUD), Data Assimilation in Discrete Event Simulations
10 Julienka Mollee (VUA), Moving forward: supporting physical activity behavior change through intelligent technology
11 Mahdi Sargolzaei (UVA), Enabling Framework for Service-oriented Collaborative Networks
12 Xixi Lu (TUE), Using behavioral context in process mining
13 Seyed Amin Tabatabaei (VUA), Computing a Sustainable Future
14 Bart Joosten (UVT), Detecting Social Signals with Spatiotemporal Gabor Filters
15 Naser Davarzani (UM), Biomarker discovery in heart failure
16 Jaebok Kim (UT), Automatic recognition of engagement and emotion in a group of children
17 Jianpeng Zhang (TUE), On Graph Sample Clustering
18 Henriette Nakad (UL), De Notaris en Private Rechtspraak
19 Minh Duc Pham (VUA), Emergent relational schemas for RDF
20 Manxia Liu (RUN), Time and Bayesian Networks
21 Aad Slootmaker (OUN), EMERGO: a generic platform for authoring and playing scenario-based serious games
22 Eric Fernandes de Mello Araújo (VUA), Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks
23 Kim Schouten (EUR), Semantics-driven Aspect-Based Sentiment Analysis
24 Jered Vroon (UT), Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots
25 Riste Gligorov (VUA), Serious Games in Audio-Visual Collections
26 Roelof Anne Jelle de Vries (UT),Theory-Based and Tailor-Made: Motivational Messages for Behavior Change Technology
27 Maikel Leemans (TUE), Hierarchical Process Mining for Scalable Software Analysis
28 Christian Willemse (UT), Social Touch Technologies: How they feel and how they make you feel
29 Yu Gu (UVT), Emotion Recognition from Mandarin Speech
30 Wouter Beek, The "K" in "semantic web" stands for "knowledge": scaling semantics to the web

---

2019 01 Rob van Eijk (UL),Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification
02 Emmanuelle Beauxis Aussalet (CWI, UU), Statistics and Visualizations for Assessing Class Size Uncertainty
03 Eduardo Gonzalez Lopez de Murillas (TUE), Process Mining on Databases: Extracting Event Data from Real Life Data Sources
04 Ridho Rahmadi (RUN), Finding stable causal structures from clinical data
05 Sebastiaan van Zelst (TUE), Process Mining with Streaming Data
06 Chris Dijkshoorn (VU), Nichesourcing for Improving Access to Linked Cultural Heritage Datasets
07 Soude Fazeli (TUD), Recommender Systems in Social Learning Platforms
08 Frits de Nijs (TUD), Resource-constrained Multi-agent Markov Decision Processes
09 Fahimeh Alizadeh Moghaddam (UVA), Self-adaptation for energy efficiency in software systems
10 Qing Chuan Ye (EUR), Multi-objective Optimization Methods for Allocation and Prediction
11 Yue Zhao (TUD), Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs

12 Jacqueline Heinerman (VU), Better Together
13 Guanliang Chen (TUD), MOOC Analytics: Learner Modeling and Content Generation
14 Daniel Davis (TUD), Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses
15 Erwin Walraven (TUD), Planning under Uncertainty in Constrained and Partially Observable Environments
16 Guangming Li (TUE), Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models
17 Ali Hurriyetoglu (RUN),Extracting actionable information from microtexts
18 Gerard Wagenaar (UU), Artefacts in Agile Team Communication
19 Vincent Koeman (TUD), Tools for Developing Cognitive Agents
20 Chide Groenouwe (UU), Fostering technically augmented human collective intelligence
21 Cong Liu (TUE), Software Data Analytics: Architectural Model Discovery and Design Pattern Detection
22 Martin van den Berg (VU),Improving IT Decisions with Enterprise Architecture
23 Qin Liu (TUD), Intelligent Control Systems: Learning, Interpreting, Verification
24 Anca Dumitrache (VU), Truth in Disagreement - Crowdsourcing Labeled Data for Natural Language Processing
25 Emiel van Miltenburg (VU), Pragmatic factors in (automatic) image description
26 Prince Singh (UT), An Integration Platform for Synchromodal Transport
27 Alessandra Antonaci (OUN), The Gamification Design Process applied to (Massive) Open Online Courses
28 Esther Kuindersma (UL), Cleared for take-off: Game-based learning to prepare airline pilots for critical situations
29 Daniel Formolo (VU), Using virtual agents for simulation and training of social skills in safety-critical circumstances
30 Vahid Yazdanpanah (UT), Multiagent Industrial Symbiosis Systems
31 Milan Jelisavcic (VU), Alive and Kicking: Baby Steps in Robotics
32 Chiara Sironi (UM), Monte-Carlo Tree Search for Artificial General Intelligence in Games
33 Anil Yaman (TUE), Evolution of Biologically Inspired Learning in Artificial Neural Networks
34 Negar Ahmadi (TUE), EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES
35 Lisa Facey-Shaw (OUN), Gamification with digital badges in learning programming
36 Kevin Ackermans (OUN), Designing Video-Enhanced Rubrics to Master Complex Skills
37 Jian Fang (TUD), Database Acceleration on FPGAs
38 Akos Kadar (OUN), Learning visually grounded and multilingual representations

2020 01 Armon Toubman (UL), Calculated Moves: Generating Air Combat Behaviour
02 Marcos de Paula Bueno (UL), Unraveling Temporal Processes using Probabilistic Graphical Models
03 Mostafa Deghani (UvA), Learning with Imperfect Supervision for Language Understanding
04 Maarten van Gompel (RUN), Context as Linguistic Bridges
05 Yulong Pei (TUE), On local and global structure mining
06 Preethu Rose Anish (UT), Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support
07 Wim van der Vegt (OUN), Towards a software architecture for reusable game components
08 Ali Mirsoleimani (UL),Structured Parallel Programming for Monte Carlo Tree Search
09 Myriam Traub (UU), Measuring Tool Bias and Improving Data Quality for Digital Humanities Research
10 Alifah Syamsiyah (TUE), In-database Preprocessing for Process Mining
11 Sepideh Mesbah (TUD), Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models
12 Ward van Breda (VU), Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment
13 Marco Virgolin (CWI), Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming
14 Mark Raasveldt (CWI/UL), Integrating Analytics with Relational Databases
15 Konstantinos Georgiadis (OUN), Smart CAT: Machine Learning for Configurable Assessments in Serious Games
16 Ilona Wilmont (RUN), Cognitive Aspects of Conceptual Modelling
17 Daniele Di Mitri (OUN), The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences
18 Georgios Methenitis (TUD), Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems
19 Guido van Capelleveen (UT), Industrial Symbiosis Recommender Systems
20 Albert Hankel (VU), Embedding Green ICT Maturity in Organisations
21 Karine da Silva Miras de Araujo (VU), Where is the robot?: Life as it could be
22 Maryam Masoud Khamis (RUN), Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar
23 Rianne Conijn (UT), The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging
24 Lenin da Nóbrega Medeiros (VUA/RUN), How are you feeling, human? Towards emotionally supportive chatbots

25 Xin Du (TUE), The Uncertainty in Exceptional Model Mining
26 Krzysztof Leszek Sadowski (UU), GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization
27 Ekaterina Muravyeva (TUD), Personal data and informed consent in an educational context
28 Bibeg Limbu (TUD), Multimodal interaction for deliberate practice: Training complex skills with augmented reality
29 Ioan Gabriel Bucur (RUN), Being Bayesian about Causal Inference
30 Bob Zadok Blok (UL), Creatief, Creatiever, Creatiefst
31 Gongjin Lan (VU), Learning better – From Baby to Better
32 Jason Rhuggenaath (TUE), Revenue management in online markets: pricing and online advertising
33 Rick Gilsing (TUE), Supporting service-dominant business model evaluation in the context of business model innovation
34 Anna Bon (MU), Intervention or Collaboration? Redesigning Information and Communication Technologies for Development
35 Siamak Farshidi (UU), Multi-Criteria Decision-Making in Software Production

---

2021 01 Francisco Xavier Dos Santos Fonseca (TUD),Location-based Games for Social Interaction in Public Space
02 Rijk Mercuur (TUD), Simulating Human Routines: Integrating Social Practice Theory in Agent-Based Models
03 Seyyed Hadi Hashemi (UVA), Modeling Users Interacting with Smart Devices
04 Ioana Jivet (OU), The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning
05 Davide Dell'Anna (UU), Data-Driven Supervision of Autonomous Systems
06 Daniel Davison (UT), "Hey robot, what do you think?" How children learn with a social robot
07 Armel Lefebvre (UU), Research data management for open science
08 Nardie Fanchamps (OU), The Influence of Sense-Reason-Act Programming on Computational Thinking
09 Cristina Zaga (UT), The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play
10 Quinten Meertens (UvA), Misclassification Bias in Statistical Learning
11 Anne van Rossum (UL), Nonparametric Bayesian Methods in Robotic Vision
12 Lei Pi (UL), External Knowledge Absorption in Chinese SMEs
13 Bob R. Schadenberg (UT), Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning
14 Negin Samaeemofrad (UL), Business Incubators: The Impact of Their Support
15 Onat Ege Adali (TU/e), Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm
16 Esam A. H. Ghaleb (UM), Bimodal emotion recognition from audio-visual cues
17 Dario Dotti (UM), Human Behavior Understanding from motion and bodily cues using deep neural networks
18 Remi Wieten (UU), Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks
19 Roberto Verdecchia (VU), Architectural Technical Debt: Identification and Management
20 Masoud Mansoury (TU/e), Understanding and Mitigating Multi-Sided Exposure Bias in Recommender Systems
21 Pedro Thiago Timbó Holanda (CWI), Progressive Indexes
22 Sihang Qiu (TUD), Conversational Crowdsourcing
23 Hugo Manuel Proença (LIACS), Robust rules for prediction and description
24 Kaijie Zhu (TUE), On Efficient Temporal Subgraph Query Processing
25 Eoin Martino Grua (VUA), The Future of E-Health is Mobile: Combining AI and Self-Adaptation to Create Adaptive E-Health Mobile Applications
26 Benno Kruit (CWI & VUA), Reading the Grid: Extending Knowledge Bases from Human-readable Tables
27 Jelte van Waterschoot (UT), Personalized and Personal Conversations: Designing Agents Who Want to Connect With You
28 Christoph Selig (UL), Understanding the Heterogeneity of Corporate Entrepreneurship Programs

---

2022 01 Judith van Stegeren (UT), Flavor text generation for role-playing video games
02 Paulo da Costa (TU/e), Data-driven Prognostics and Logistics Optimisation: A Deep Learning Journey
03 Ali el Hassouni (VUA), A Model A Day Keeps The Doctor Away: Reinforcement Learning For Personalized Healthcare
04 Ünal Aksu (UU), A Cross-Organizational Process Mining Framework
05 Shiwei Liu (TU/e), Sparse Neural Network Training with In-Time Over-Parameterization
06 Reza Refaei Afshar (TU/e), Machine Learning for Ad Publishers in Real Time Bidding
07 Sambit Praharaj (OU), Measuring the Unmeasurable? Towards Automatic Co-located Collaboration Analytics
08 Maikel L. van Eck (TU/e), Process Mining for Smart Product Design
09 Oana Andreea Inel (VUA), Understanding Events: A Diversity-driven Human-Machine Approach

10 Felipe Moraes Gomes (TUD), Examining the Effectiveness of Collaborative Search Engines
11 Mirjam de Haas (UT), Staying engaged in child-robot interaction, a quantitative approach to studying preschoolers' engagement with robots and tasks during second-language tutoring
12 Guanyi Chen (UU), Computational Generation of Chinese Noun Phrases
13 Xander Wilcke (VUA), Machine Learning on Multimodal Knowledge Graphs: Opportunities, Challenges, and Methods for Learning on Real-World Heterogeneous and Spatially-Oriented Knowledge
14 Michiel Overeem (UU), Evolution of Low-Code Platforms
15 Jelmer Jan Koorn (UU), Work in Process: Unearthing Meaning using Process Mining
16 Pieter Gijsbers (TU/e), Systems for AutoML Research
17 Laura van der Lubbe (VUA), Empowering vulnerable people with serious games and gamification
18 Paris Mavromoustakos Blom (TiU), Player Affect Modelling and Video Game Personalisation
19 Bilge Yigit Ozkan (UU), Cybersecurity Maturity Assessment and Standardisation
20 Fakhra Jabeen (VUA), Dark Side of the Digital Media - Computational Analysis of Negative Human Behaviors on Social Media
21 Seethu Mariyam Christopher (UM), Intelligent Toys for Physical and Cognitive Assessments
22 Alexandra Sierra Rativa (TiU), Virtual Character Design and its potential to foster Empathy, Immersion, and Collaboration Skills in Video Games and Virtual Reality Simulations
23 Ilir Kola (TUD), Enabling Social Situation Awareness in Support Agents
24 Samaneh Heidari (UU), Agents with Social Norms and Values - A framework for agent based social simulations with social norms and personal values
25 Anna L.D. Latour (LU), Optimal decision-making under constraints and uncertainty
26 Anne Dirkson (LU), Knowledge Discovery from Patient Forums: Gaining novel medical insights from patient experiences
27 Christos Athanasiadis (UM), Emotion-aware cross-modal domain adaptation in video sequences
28 Onuralp Ulusoy (UU), Privacy in Collaborative Systems
29 Jan Kolkmeier (UT), From Head Transform to Mind Transplant: Social Interactions in Mixed Reality
30 Dean De Leo (CWI), Analysis of Dynamic Graphs on Sparse Arrays
31 Konstantinos Traganos (TU/e), Tackling Complexity in Smart Manufacturing with Advanced Manufacturing Process Management
32 Cezara Pastrav (UU), Social simulation for socio-ecological systems
33 Brinn Hekkelman (CWI/TUD), Fair Mechanisms for Smart Grid Congestion Management
34 Nimat Ullah (VUA), Mind Your Behaviour: Computational Modelling of Emotion & Desire Regulation for Behaviour Change
35 Mike E.U. Ligthart (VUA), Shaping the Child-Robot Relationship: Interaction Design Patterns for a Sustainable Interaction

---

2023 01 Bojan Simoski (VUA), Untangling the Puzzle of Digital Health Interventions
02 Mariana Rachel Dias da Silva (TiU), Grounded or in flight? What our bodies can tell us about the whereabouts of our thoughts
03 Shabnam Najafian (TUD), User Modeling for Privacy-preserving Explanations in Group Recommendations
04 Gineke Wiggers (UL), The Relevance of Impact: bibliometric-enhanced legal information retrieval
05 Anton Bouter (CWI), Optimal Mixing Evolutionary Algorithms for Large-Scale Real-Valued Optimization, Including Real-World Medical Applications
06 António Pereira Barata (UL), Reliable and Fair Machine Learning for Risk Assessment
07 Tianjin Huang (TU/e), The Roles of Adversarial Examples on Trustworthiness of Deep Learning
08 Lu Yin (TU/e), Knowledge Elicitation using Psychometric Learning
09 Xu Wang (VUA), Scientific Dataset Recommendation with Semantic Techniques
10 Dennis J.N.J. Soemers (UM), Learning State-Action Features for General Game Playing
11 Fawad Taj (VUA), Towards Motivating Machines: Computational Modeling of the Mechanism of Actions for Effective Digital Health Behavior Change Applications
12 Tessel Bogaard (VUA), Using Metadata to Understand Search Behavior in Digital Libraries
13 Injy Sarhan (UU), Open Information Extraction for Knowledge Representation
14 Selma Čaušević (TUD), Energy resilience through self-organization
15 Alvaro Henrique Chaim Correia (TU/e), Insights on Learning Tractable Probabilistic Graphical Models
16 Peter Blomsma (TiU), Building Embodied Conversational Agents: Observations on human nonverbal behaviour as a resource for the development of artificial characters
17 Meike Nauta (UT), Explainable AI and Interpretable Computer Vision – From Oversight to Insight
18 Gustavo Penha (TUD), Designing and Diagnosing Models for Conversational Search and Recommendation
19 George Aalbers (TiU), Digital Traces of the Mind: Using Smartphones to Capture Signals of Well-Being in Individuals
20 Arkadiy Dushatskiy (TUD), Expensive Optimization with Model-Based Evolutionary Algorithms applied to Medical Image Segmentation using Deep Learning

21 Gerrit Jan de Bruin (UL), Network Analysis Methods for Smart Inspection in the Transport Domain
22 Alireza Shojaifar (UU), Volitional Cybersecurity
23 Theo Theunissen (UU), Documentation in Continuous Software Development
24 Agathe Balayn (TUD), Practices Towards Hazardous Failure Diagnosis in Machine Learning
25 Jurian Baas (UU), Entity Resolution on Historical Knowledge Graphs
26 Loek Tonnaer (TU/e), Linearly Symmetry-Based Disentangled Representations and their Out-of-Distribution Behaviour
27 Ghada Sokar (TU/e), Learning Continually Under Changing Data Distributions
28 Floris den Hengst (VUA), Learning to Behave: Reinforcement Learning in Human Contexts
29 Tim Draws (TUD), Understanding Viewpoint Biases in Web Search Results

2024 01 Daphne Miedema (TU/e), On Learning SQL: Disentangling concepts in data systems education
02 Emile van Krieken (VUA), Optimisation in Neurosymbolic Learning Systems
03 Feri Wijayanto (RUN), Automated Model Selection for Rasch and Mediation Analysis
04 Mike Huisman (UL), Understanding Deep Meta-Learning
05 Yiyong Gou (UM), Aerial Robotic Operations: Multi-environment Cooperative Inspection & Construction Crack Autonomous Repair
06 Azqa Nadeem (TUD), Understanding Adversary Behavior via XAI: Leveraging Sequence Clustering to Extract Threat Intelligence
07 Parisa Shayan (TiU), Modeling User Behavior in Learning Management Systems
08 Xin Zhou (UvA), From Empowering to Motivating: Enhancing Policy Enforcement through Process Design and Incentive Implementation
09 Giso Dal (UT), Probabilistic Inference Using Partitioned Bayesian Networks
10 Cristina-Iulia Bucur (VUA), Linkflows: Towards Genuine Semantic Publishing in Science
11 withdrawn
12 Peide Zhu (TUD), Towards Robust Automatic Question Generation For Learning
13 Enrico Liscio (TUD), Context-Specific Value Inference via Hybrid Intelligence
14 Larissa Capobianco Shimomura (TU/e), On Graph Generating Dependencies and their Applications in Data Profiling
15 Ting Liu (VUA), A Gut Feeling: Biomedical Knowledge Graphs for Interrelating the Gut Microbiome and Mental Health
16 Arthur Barbosa Câmara (TUD), Designing Search-as-Learning Systems
17 Razieh Alidoosti (VUA), Ethics-aware Software Architecture Design
18 Laurens Stoop (UU), Data Driven Understanding of Energy-Meteorological Variability and its Impact on Energy System Operations
19 Azadeh Mozafari Mehr (TU/e), Multi-perspective Conformance Checking: Identifying and Understanding Patterns of Anomalous Behavior
20 Ritsart Anne Plantenga (UL), Omgang met Regels
21 Federica Vinella (UU), Crowdsourcing User-Centered Teams
22 Zeynep Ozturk Yurt (TU/e), Beyond Routine: Extending BPM for Knowledge-Intensive Processes with Controllable Dynamic Contexts
23 Jie Luo (VUA), Lamarck's Revenge: Inheritance of Learned Traits Improves Robot Evolution
24 Nirmal Roy (TUD), Exploring the Effects of Interactive Interfaces on User Search Behaviour
25 Alisa Rieger (TUD), Striving for Responsible Opinion Formation in Web Search on Debated Topics