From Potential to Practice
Intellectual Humility During Search on Debated Topics

Rieger, Alisa; Bredius, Frank; Theune, Mariët; Pera, Maria Soledad

DOI
10.1145/3627508.3638306

Publication date
2024

Document Version
Final published version

Published in
CHIIR 2024 - Proceedings of the 2024 Conference on Human Information Interaction and Retrieval

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

Alisa Rieger
a.rieger@tudelft.nl
Delft University of Technology
Delft, Netherlands

Frank Bredius
F.Bredius@student.tudelft.nl
Delft University of Technology
Delft, Netherlands

Mariët Theune
m.theune@utwente.nl
University of Twente
Enschede, Netherlands

Maria Soledad Pera
M.S.Pera@tudelft.nl
Delft University of Technology
Delft, Netherlands

ABSTRACT

An essential characteristic for unbiased and diligent information-seeking that can enable informed opinion formation and decision-making is *intellectual humility* (IH), the awareness of the limitations of one's knowledge and opinions. While researchers have recognized the potential to boost IH in individuals, the effect of such interventions on their search behavior, along with the broader significance of IH in the context of web search on debated topics remains unexplored. In this paper, we present the results of a preregistered user study ($N = 299$) that we conducted to (1) test the effect of three interventions that boost self-reported IH on opinionated individuals' search behavior and (2) explore the role of IH in the search process of opinionated individuals more broadly. IH-boosting interventions did not affect search behavior; we attribute this to the high familiarity of the search environment, prompting searchers to default to their usual search behavior. Still, explorations of the role of IH in the search process indicate that IH and IH-related search intentions should be considered as relevant factors in the pursuit of supporting unbiased and diligent search on debated topics. Based on our exploratory findings, we argue that future research should investigate interventions that are more directly integrated into the search process, as well as such that combine boosting IH with encouraging searchers to approach the search task in an IH-driven way and promoting transparency for appropriate reliance on the search system and ranking.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; *User centered design*; • **Information systems** → **Search interfaces**.

KEYWORDS

Web Search, Intellectual Humility, Boosting, Debated Topics, Confirmation Bias

CHIIR '24, March 10–14, 2024, Sheffield, United Kingdom

ACM ISBN 979-8-4007-0434-5/24/03
https://doi.org/10.1145/3627508.3638306

ACM Reference Format:
Alisa Rieger, Frank Bredius, Mariët Theune, and Maria Soledad Pera. 2024. From Potential to Practice: Intellectual Humility during Search on Debated Topics. In *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '24), March 10–14, 2024, Sheffield, United Kingdom.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3627508.3638306

1 INTRODUCTION

It has become increasingly common for individuals to rely on web search engines as their starting point when seeking information [9, 16, 19, 53]. Searchers tend to deem their own search interactions as unbiased and trust search engines, perceiving highly-ranked search results as relevant and accurate [9, 20, 22]. According to Smith and Rieh [54], this trust leads searchers to believe that they do not need to invest cognitive effort in the search process. High reliance on search engines can lead to efficient task completion for simple lookup search tasks, e.g., to retrieve facts and for question answering. However, people also use search engines for non-arbitrary, i.e., complex, information needs, such as learning and investigation [39]. Many complex search tasks inherently require cognitive effort and diligence to identify knowledge gaps and actively explore, compare, and critically evaluate different resources [39, 53].

Over-relying on the search system and conducting online inquiries with low cognitive effort and diligence, impeding knowledge gain, is particularly problematic for consequential searches, such as searching for resources on **debated topics**. When it comes to these often controversial matters, individuals might carry out such searches because they want to form an informed opinion—*whether the state should provide a universal basic income*—or make an informed decision—*whom to vote for*. In theory, in both these cases, searchers would, ideally, actively engage with diverse viewpoints, explore a variety of resources, and verify the accuracy of the information they encounter [30, 44]. In practice, however, searching for information on debated topics can prove challenging because debated topics are multifaceted and complex, with perspectives that are linked to different values or interests, some of which might threaten an individual's own values and interests [26, 45]. This can lead to emotionally charged and biased search behavior, for instance, when individuals prioritize information that aligns with their preexisting attitudes while dismissing or discounting arguments that challenge their views *(confirmation bias)* [4, 43, 62].

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, poses an obstacle to informed opinion formation or decision-making [41]. This is compounded by the fact that, in general, searchers' information seeking habits are known to be shaped by various individual characteristics [7, 42, 45, 60], 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 [10, 45]. Individuals with high IH generally have a high motivation to seek information and gain knowledge [34, 46]. They tend to spend more time learning about attitude-opposing arguments and can better identify the strength of different arguments [6], 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 [33], reading about the benefits of IH [47], or reading about the plasticity of intelligence [46], 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 [45, 48]. 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 [59] and thus in greater need of support for unbiased search. Guided by three preregistered and one exploratory research questions, 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 3 (*prime, remind, reinforce*, see Figure 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?
- **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?

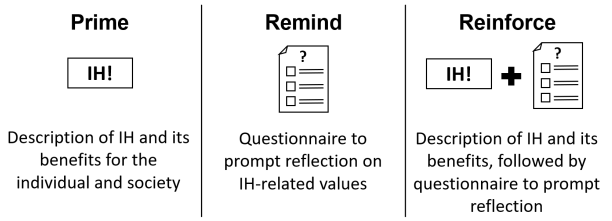

Figure 1: Boosting interventions inspired by the concept of Intellectual Humility (IH).

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 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 [38].

Contributions. This paper advances 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 data set with questionnaire responses and behavioral data from search logs publicly available.[1]

2 RELATED WORK

Most search engines lack support for the activities searchers need to carry out to satisfy complex information needs [36, 39, 53, 54]. For instance, mainstream search engines may fail to support diverse search intentions, information-seeking strategies, and transitions between them [36, 53]. 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 [31, 41]. 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 [51, 53, 54].

To *empower individuals* to navigate online environments, Lorenz-Spreen et al. [38] 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) [8, 56]. In contrast, boosting interventions aim at fostering user competencies that facilitate navigating online environments and, unlike nudging, offer the advantage of upholding user autonomy, as well as remaining effective over an extended period of time [25, 38]. 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 [49], displaying labels that indicate the stance of the search results [63], or tag clouds that reveal experts evaluations [52] were found to reduce confirmation bias, while query priming was found to promote diligent search behavior and increased exploration [64]. Researchers have also investigated *argument retrieval* systems that can facilitate web search on debated topics by directly presenting distinct arguments retrieved from search results [1, 5, 61]. Boosting interventions, promising to empower users to navigate other online challenges, such as microtargeting [37], 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 [7, 42, 45, 60], 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 [59]), and others are more stable (e.g., the need for cognition—an individual's general tendency to organize their experience meaningfully [7, 60]). 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 [59], and individuals with a low compared to high *need for cognition* were observed to be less diligent searchers [60]. 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 [10, 45]. 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) [3, 27, 45]. 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 [34, 47], as well as little engagement with opposing viewpoints [6]. Looking at societal challenges arising from these information seeking patterns, high IH was linked to reduced hostility towards individuals with opposing views [55], decreased affective polarization [6], and diminished susceptibility to misinformation [29]. In light of these observations, researchers see potential in interventions that boost IH to function as an antidote to such epistemic societal challenges [21, 45]. While simple approaches that effectively boost self-reported IH have indeed been identified [33, 46, 47], their effect on real-world information behavior, and web search on debated topics, in particular, has yet to be explored.

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 could foster web users' cognitive competencies (for an overview of different digital boosting approaches, see https://www.scienceofboosting.org/tag/digital/). 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 an approach similar to [38] and considered alternatives of varying complexity. This resulted in three boosting interventions (see Figure 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* [3];

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

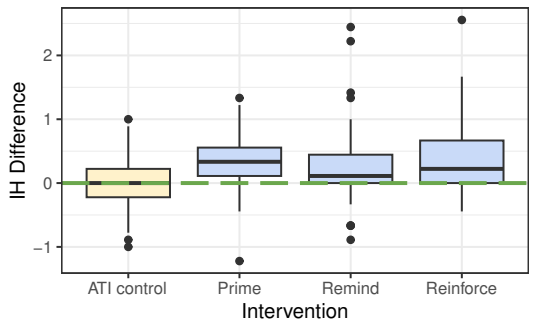

Figure 2: Prestudy: IH difference per intervention condition. Difference in IH levels before and after participants' exposure to one of the considered interventions.

(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* [3].

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* (https://www.prolific.com), 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. [11] 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) [27]. 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, 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 2). Based on these findings, we tested the effect of all three boosting interventions on search behavior in the main user study.

4 USER STUDY METHODOLOGY

To investigate the three preregistered RQs (see Section 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[2]):** 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 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.[3]

4.1 Procedure

We recruited participants via *Prolific* and used *Qualtrics* (https://www.qualtrics.com) for pre- and post-search questionnaires. We collected the data for this user study with the following procedure (see Figure 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

[2] 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.
[3] 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.

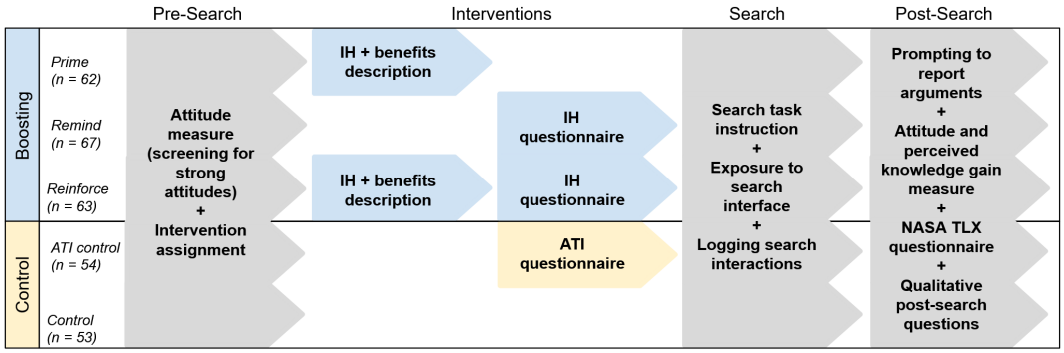

Figure 3: Procedure of the user study per intervention condition.

conditions (control, ATI control, prime, remind, reinforce, see Figure 3). We tested two control conditions, one without any intervention and one in which participants were asked to fill 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 form the data-set by Draws et al. [11]. 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 [40].
- **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., [13, 15, 50]). 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 [23], 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.

4.2 Variables

In Table 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 rank clicked, dwell time, task completion time, cumulative clicks)[4], as well as to determine differences in search behavior across the five interventions (**RQ3, H3**). To investigate how IH factor 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 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.

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- β) = 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% above 65 years old. Participation was rewarded with £2.30 (mean = £9.32/h).

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

Table 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.

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 (§ 5.1). 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 (§ 5.2).

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 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 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 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, we could not corroborate that any of the boosting interventions impact search diligence since we did not find evidence for H2a-d.

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.

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 ($r = -0.09, p = 0.29$) 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 ($r = 0.25, p = 0.004$), where searchers with higher IH reported higher knowledge gain.

Table 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

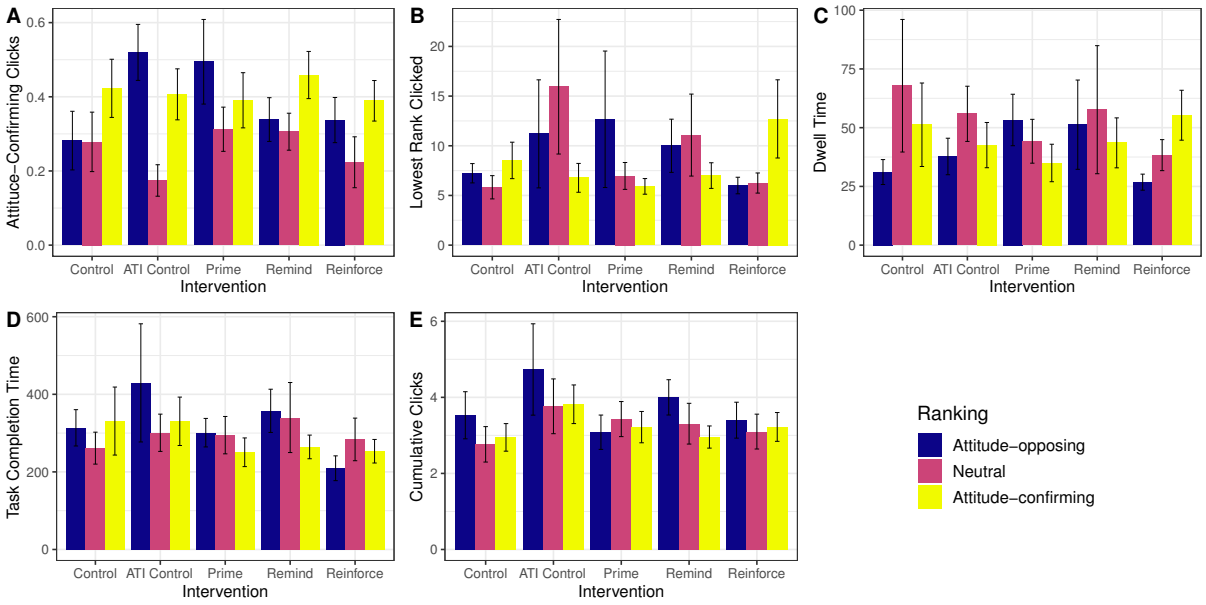

Figure 4: Search Behavior per Intervention and Ranking. Mean (A) proportion of attitude-confirming clicks, (B) lowest rank clicked, (C) dwell time, (D) task completion time, and (E) number of cumulative clicks per intervention (control, ATI control, prime, remind, reinforce) and ranking condition with 95% confidence interval.

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

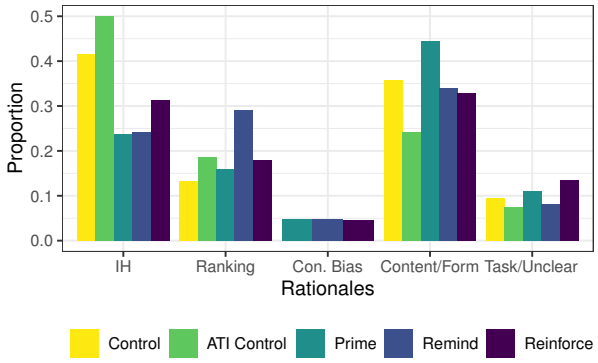

Figure 5: Proportion of participants in each intervention condition who reported a given rationale.

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.*

We evaluated the proportion of participants per intervention condition who reported each rationale (see Figure 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 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 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 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 across 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 ($r = -0.18, p = .042$) and temporal demand ($r = -0.21, p = .014$). Further, we observed moderate correlations to performance ($r = 0.4, p < .001$) and frustration ($r = -0.44, p < .001$). 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.

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 (§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 [6, 34, 46]). 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 [33, 46, 47], 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

Table 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

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 [2, 14]. 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 [54], 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 [25, 38], 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 [6, 21, 34, 46], 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 [35] 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* [58], or *attitude importance* [26] moderate the effects of various interventions to support unbiased and diligentsearch 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.

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. Thus, Promoting 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 [38, 54]. 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. [12], 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. [17] 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 [57, 62] that are predominantly under the control of private industry [18] 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 [31, 32, 65].

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. [24], the role of IH in various information settings that extend beyond web search warrants investigation.

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. [13]. Similar to most studies 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 [28], 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 CONCLUSION

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. 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.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Yamen Ajjour, Pavel Braslavski, Alexander Bondarenko, and Benno Stein. 2022. 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)*. Association for Computing Machinery, New York, NY, USA, 2393–2399. https://doi.org/10.1145/3477495.3531864

[2] Marwah Alaofi, Luke Gallagher, Dana Mckay, Lauren L. Saling, Mark Sanderson, Falk Scholer, Damiano Spina, and Ryen W. White. 2022. Where Do Queries Come From?. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 2850–2862. https://doi.org/10.1145/3477495.3531711

[3] Mark Alfano, Kathryn Iurino, Paul Stey, Brian Robinson, Markus Christen, Feng Yu, and Daniel Lapsley. 2017. Development and Validation of a Multi-Dimensional Measure of Intellectual Humility. *PLOS ONE* 12 (2017), e0182950. https://doi.org/10.1371/journal.pone.0182950

[4] Leif Azzopardi. 2021. Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. ACM, Canberra ACT Australia, 27–37. https://doi.org/10.1145/3406522.3446023

[5] 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. 2022. Overview of Touché 2022: Argument Retrieval. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction (Lecture Notes in Computer Science)*, 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 (Eds.). Springer International Publishing, Cham, 311–336. https://doi.org/10.1007/978-3-031-13643-6_21

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

[7] John T. Cacioppo, Richard E. Petty, and Chuan Feng Kao. 1984. The Efficient Assessment of Need for Cognition. *Journal of Personality Assessment* 48 (1984), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

[8] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 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*. Association for Computing Machinery, New York, NY, USA, 1–15.

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

[10] Don E. Davis, Kenneth Rice, Stacey McElroy, Cirleen DeBlaere, Elise Choe, Daryl R. Van Tongeren, and Joshua N. Hook. 2016. Distinguishing Intellectual Humility and General Humility. *The Journal of Positive Psychology* 11 (2016), 215–224. https://doi.org/10.1080/17439760.2015.1048818

[11] Tim Draws, Alisa Rieger, Oana Inel, Ujwal Gadiraju, and Nava Tintarev. 2021. A checklist to combat cognitive biases in crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 9. 48–59.

[12] Tim Draws, Nirmal Roy, Oana Inel, Alisa Rieger, Rishav Hada, Mehmet Orcun Yalcin, Benjamin Timmermans, and Nava Tintarev. 2023. Viewpoint Diversity in Search Results. In *Advances in Information Retrieval (Lecture Notes in Computer Science)*, Jaap Kamps, Lorraine Goeuriot, Fabio Crestani, Maria Maistro, Hideo Joho, Brian Davis, Cathal Gurrin, Udo Kruschwitz, and Annalina Caputo (Eds.). Springer Nature Switzerland, Cham, 279–297. https://doi.org/10.1007/978-3-031-28244-7_18

[13] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. 2021. This Is Not What We Ordered: Exploring Why Biased Search Result Rankings Affect User Attitudes on Debated Topics. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Virtual Event Canada, 295–305. https://doi.org/10.1145/3404835.3462851

[14] Karl Duncker. 1945. On Problem-Solving. *Psychological Monographs* 58 (1945), i–113. https://doi.org/10.1037/h0093599

[15] Robert Epstein and Ronald E. Robertson. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences* 112, 33 (Aug. 2015), E4512–E4521. https://doi.org/10.1073/pnas.1419828112

[16] World Leaders in Research-Based User Experience. 2020. How Information-Seeking Behavior Has Changed in 22 Years. https://www.nngroup.com/articles/information-seeking-behavior-changes/.

[17] Isabelle Freiling, Nicole M. Krause, and Dietram A. Scheufele. 2023. Science and Ethics of "Curing" Misinformation. *AMA Journal of Ethics* 25 (2023), 228–237. https://doi.org/10.1001/amajethics.2023.228

[18] Maria Garrido and Stephen Wyber. 2017. *Development and Access to Information 2017*. Report. International Federation of Library Associations and Institutions.

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

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

[21] Tim Gorichanaz. 2022. Relating Information Seeking and Use to Intellectual Humility. *Journal of the Association for Information Science and Technology* 73 (2022), 643–654. https://doi.org/10.1002/asi.24567

[22] Eszter Hargittai, Lindsay Fullerton, Ericka Menchen-Trevino, and Kristin Yates Thomas. 2010. Trust online: Young adults' evaluation of web content. *International journal of communication* 4 (2010), 27.

[23] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[24] Amelia Hassoun, Ian Beacock, Sunny Consolvo, Beth Goldberg, Patrick Gage Kelley, and Daniel M. Russell. 2023. 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)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3544548.3581328

[25] Ralph Hertwig and Till Grüne-Yanoff. 2017. Nudging and Boosting: Steering or Empowering Good Decisions. *Perspectives on Psychological Science* 12 (2017), 973–986. https://doi.org/10.1177/1745691617702496

[26] Lauren C. Howe and Jon A. Krosnick. 2017. Attitude Strength. *Annual Review of Psychology* 68 (2017), 327–351. https://doi.org/10.1146/annurev-psych-122414-033600

[27] Rick H. Hoyle, Erin K. Davisson, Kate J. Diebels, and Mark R. Leary. 2016. Holding Specific Views with Humility: Conceptualization and Measurement of Specific Intellectual Humility. *Personality and Individual Differences* 97 (2016), 165–172. https://doi.org/10.1016/j.paid.2016.03.043

[28] Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The Future of Crowd Work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1301–1318. https://doi.org/10.1145/2441776.2441923

[29] Jonah Koetke, Karina Schumann, and Tenelle Porter. 2022. Intellectual Humility Predicts Scrutiny of COVID-19 Misinformation. *Social Psychological and Personality Science* 13 (2022), 277–284. https://doi.org/10.1177/1948550620988242

[30] Hilary Kornblith. 1983. Justified Belief and Epistemically Responsible Action. *The Philosophical Review* 92 (1983), 33–48. https://doi.org/10.2307/2184520

[31] Anastasia Kozyreva, Stephan Lewandowsky, and Ralph Hertwig. 2020. Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools. *Psychological Science in the Public Interest* 21 (2020), 103–156. https://doi.org/10.1177/1529100620946707

[32] Anastasia Kozyreva, Sam Wineburg, Stephan Lewandowsky, and Ralph Hertwig. 2023. Critical Ignoring as a Core Competence for Digital Citizens. *Current Directions in Psychological Science* 32 (2023), 81–88. https://doi.org/10.1177/09637214221121570

[33] Ethan Kross and Igor Grossmann. 2012. Boosting Wisdom: Distance from the Self Enhances Wise Reasoning, Attitudes, and Behavior. *Journal of Experimental Psychology: General* 141 (2012), 43–48. https://doi.org/10.1037/a0024158

[34] Elizabeth J. Krumrei-Mancuso, Megan C. Haggard, Jordan P. LaBouff, and Wade C. Rowatt. 2020. Links between Intellectual Humility and Acquiring Knowledge. *The Journal of Positive Psychology* 15 (2020), 155–170. https://doi.org/10.1080/17439760.2019.1579359

[35] Elizabeth J. Krumrei-Mancuso and Brian Newman. 2020. Intellectual Humility in the Sociopolitical Domain. *Self and Identity* 19 (2020), 989–1016. https://doi.org/10.1080/15298868.2020.1714711

[36] Jiqun Liu, Matthew Mitsui, Nicholas J. Belkin, and Chirag Shah. 2019. 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)*. Association for Computing Machinery, New York, NY, USA, 123–132. https://doi.org/10.1145/3295750.3298922

[37] Philipp Lorenz-Spreen, Michael Geers, Thorsten Pachur, Ralph Hertwig, Stephan Lewandowsky, and Stefan M. Herzog. 2021. Boosting People's Ability to Detect Microtargeted Advertising. *Scientific Reports* 11 (2021), 15541. https://doi.org/10.1038/s41598-021-94796-z

[38] Philipp Lorenz-Spreen, Stephan Lewandowsky, Cass R. Sunstein, and Ralph Hertwig. 2020. How Behavioural Sciences Can Promote Truth, Autonomy and Democratic Discourse Online. *Nature Human Behaviour* 4 (2020), 1102–1109. https://doi.org/10.1038/s41562-020-0889-7

[39] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.

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

[41] Boaz Miller and Isaac Record. 2013. Justified Belief in the Digital Age: On the Epistemic Implications of Secret Internet Technologies. *Episteme* 10 (2013), 117–134. https://doi.org/10.1017/epi.2013.11

[42] Julia A. Minson, Frances S. Chen, and Catherine H. Tinsley. 2019. Why Won't You Listen to Me? Measuring Receptiveness to Opposing Views. *Management Science* (2019). https://doi.org/10.1287/mnsc.2019.3362

[43] Raymond S Nickerson. 1998. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. (1998), 46.

[44] Rik Peels. 2016. *Responsible Belief: A Theory in Ethics and Epistemology*. Oxford University Press.

[45] Tenelle Porter, Abdo Elnakouri, Ethan A. Meyers, Takuya Shibayama, Eranda Jayawickreme, and Igor Grossmann. 2022. Predictors and Consequences of Intellectual Humility. *Nature Reviews Psychology* 1 (2022), 524–536. https://doi.org/10.1038/s44159-022-00081-9

[46] Tenelle Porter and Karina Schumann. 2018. Intellectual Humility and Openness to the Opposing View. *Self and Identity* 17 (2018), 139–162. https://doi.org/10.1080/15298868.2017.1361861

[47] Tenelle Porter, Karina Schumann, Diana Selmeczy, and Kali Trzesniewski. 2020. Intellectual Humility Predicts Mastery Behaviors When Learning. *Learning and Individual Differences* 80 (2020), 101888. https://doi.org/10.1016/j.lindif.2020.101888

[48] Alisa Rieger, Frank Bredius, Nava Tintarev, and Maria Soledad Pera. 2023. 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)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3544549.3585693

[49] Alisa Rieger, Tim Draws, Mariët Theune, and Nava Tintarev. 2021. This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias. In *Proceedings of the 32st ACM Conference on Hypertext and Social Media*. ACM, Virtual Event USA, 189–199. https://doi.org/10.1145/3465336.3475101

[50] Alisa Rieger, Tim Draws, Mariët Theune, and Nava Tintarev. 2023. Nudges to Mitigate Confirmation Bias during Web Search on Debated Topics: Support vs. Manipulation. *ACM Transactions on the Web* (2023). https://doi.org/10.1145/3635034

[51] Tuukka Ruotsalo, Giulio Jacucci, Petri Myllymäki, and Samuel Kaski. 2015. Interactive Intent Modeling: Information Discovery beyond Search. *Commun. ACM* 58 (2015), 86–92. https://doi.org/10.1145/2656334

[52] Stefan Schweiger, Aileen Oeberst, and Ulrike Cress. 2014. Confirmation Bias in Web-Based Search: A Randomized Online Study on the Effects of Expert Information and Social Tags on Information Search and Evaluation. *Journal of Medical Internet Research* 16 (2014), e3044. https://doi.org/10.2196/jmir.3044

[53] Chirag Shah and Emily M. Bender. 2022. Situating Search. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 221–232. https://doi.org/10.1145/3498366.3505816

[54] Catherine L. Smith and Soo Young Rieh. 2019. Knowledge-Context in Search Systems: Toward Information-Literate Actions. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR '19)*. Association for Computing Machinery, New York, NY, USA, 55–62. https://doi.org/10.1145/3295750.3298940

[55] Matthew L. Stanley, Alyssa H. Sinclair, and Paul Seli. 2020. Intellectual Humility and Perceptions of Political Opponents. *Journal of Personality* 88 (2020), 1196–1216. https://doi.org/10.1111/jopy.12566

[56] Richard H Thaler and Cass R Sunstein. 2008. Nudge: improving decisions about health. *Wealth, and Happiness* 6 (2008), 14–38.

[57] Kjerstin Thorson and Chris Wells. 2016. Curated Flows: A Framework for Mapping Media Exposure in the Digital Age. *Communication Theory* 26 (2016), 309–328. https://doi.org/10.1111/comt.12087

[58] Zakary L. Tormala and Derek D. Rucker. 2018. Attitude Certainty: Antecedents, Consequences, and New Directions. *Consumer Psychology Review* 1 (2018), 72–89. https://doi.org/10.1002/arcp.1004

[59] Johan L. H. van Strien, Yvonne Kammerer, Saskia Brand-Gruwel, and Henny P. A. Boshuizen. 2016. How Attitude Strength Biases Information Processing and Evaluation on the Web. *Computers in Human Behavior* 60 (2016), 245–252. https://doi.org/10.1016/j.chb.2016.02.057

[60] Bas Verplanken, Pieter T Hazenberg, and Grace R Palenéwen. 1992. Need for Cognition and External Information Search Effort. *Journal of Research in Personality* 26 (1992), 128–136. https://doi.org/10.1016/0092-6566(92)90049-A

[61] Henning Wachsmuth, Martin Potthast, Khalid Al-Khatib, Yamen Ajjour, Jana Puschmann, Jiani Qu, Jonas Dorsch, Viorel Morari, Janek Bevendorff, and Benno Stein. 2017. Building an Argument Search Engine for the Web. In *Proceedings of the 4th Workshop on Argument Mining*. Association for Computational Linguistics, Copenhagen, Denmark, 49–59. https://doi.org/10.18653/v1/W17-5106

[62] Ryen White. 2013. Beliefs and Biases in Web Search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13)*. Association for Computing Machinery, New York, NY, USA, 3–12. https://doi.org/10.1145/2484028.2484053

[63] Z. Wu, T. Draws, F. Cau, F. Barile, A. Rieger, and N. Tintarev. 2023. Explaining Search Result Stances to Opinionated People. https://doi.org/10.48550/arXiv.2309.08460 arXiv:2309.08460 [cs]

[64] Yusuke Yamamoto and Takehiro Yamamoto. 2018. Query Priming for Promoting Critical Thinking in Web Search. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. Association for Computing Machinery, New York, NY, USA, 12–21. https://doi.org/10.1145/3176349.3176377

[65] Steven Zimmerman, Stefan M. Herzog, David Elsweiler, Jon Chamberlain, and Udo Kruschwitz. 2020. 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*, Ingo Frommholz, Haiming Liu, and Massimo Melucci (Eds.), Vol. 2741. CEUR Workshop Proceedings, Xi'an, China (online)., 30–46.