Assessing Susceptibility Factors of Confirmation Bias in News Feed Reading

Nattapat Boonprakong
School of Computing and Information Systems
University of Melbourne
Parkville, Victoria, Australia
nboonprakong@student.unimelb.edu.au

Saumya Pareek
School of Computing and Information Systems
University of Melbourne
Melbourne, Victoria, Australia
spareek@student.unimelb.edu.au

Benjamin Tag
School of Computer Science and Engineering
University of New South Wales
Sydney, New South Wales, Australia
benjamin.tag@unsw.edu.au

Jorge Goncalves
School of Computing and Information Systems
University of Melbourne
Melbourne, Australia
jorge.goncalves@unimelb.edu.au

Tilman Dingler
Industrial Design Engineering
Delft University of Technology
Delft, Netherlands
t.dingler@tudelft.nl

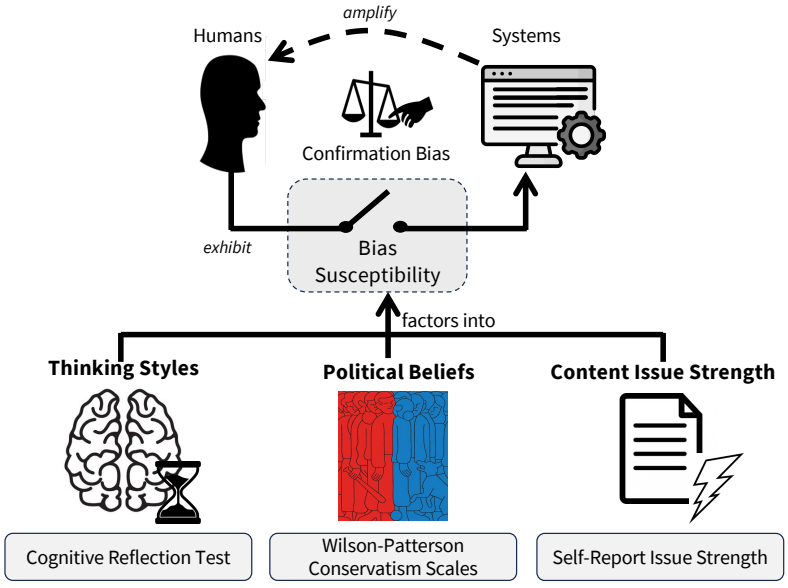


Figure 1: We investigated the human factors influencing susceptibility to confirmation bias in news feed reading and identified the individual's thinking styles, the strength of political beliefs, and the content's perceived issue strength as key contributors.

Abstract

Individuals tend to apply preferences and beliefs as heuristics to effectively sift through the sheer amount of information available online. Such tendencies, however, often result in cognitive biases, which can skew judgment and open doors for manipulation. In this work, we investigate how individual and contextual factors lead to instances of confirmation bias when seeking, evaluating, and recalling polarising information. We conducted a lab study, in which we exposed participants to opinions on controversial issues through a Twitter-like news feed. We found that low-effortful thinking, strong political beliefs, and content conveying a strong issue amplify the occurrences of confirmation bias, leading to skewed information processing and recall. We discuss how the adverse effects of confirmation bias can be mitigated by taking bias-susceptibility into account. Specifically, social media platforms could aim to reduce strong expressions and integrate media literacy-building mechanisms, as low-effortful thinking styles and strong political beliefs render individuals especially susceptible to cognitive biases.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713873

Keywords

cognitive bias, confirmation bias, individual difference, recall, social media

ACM Reference Format:

Nattapat Boonprakong, Saumya Pareek, Benjamin Tag, Jorge Goncalves, and Tilman Dingler. 2025. Assessing Susceptibility Factors of Confirmation Bias in News Feed Reading. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3706598.3713873

1 Introduction

Given the sheer amount of information available online, individuals apply **cognitive biases** as their "rule of thumb" to effectively skim through this information [5]. However, cognitive biases often skew our judgment and prompt us to give up analytical thinking [105]. These biases, therefore, can be harmful as they open doors for manipulation. Misinformation and conspiracy theories, for example, tend to trigger our cognitive biases to create more engagement about controversial, divisive issues that touch on people's preexisting beliefs (*e.g.*, politics and human rights) [18, 35]. Social engineering attacks also try to tap into individuals' cognitive biases to steer their behaviours [14]. In the Cambridge Analytica scandal [10], for example, people's personal tendencies on social media were (mis-)used to target their cognitive vulnerabilities and subsequently sway their opinion-making all without their awareness.

When consuming online information, individuals often rely on their existing preferences and beliefs as cognitive strategies – shaped by their previous experience of the world [56, 141] – to effectively process information presented to them. However, this often results in cognitive biases, which prompt individuals to see only what they want to see without carefully inspecting the content piece [6, 117]: for instance, *confirmation bias* increases people's tendency to predominantly seek, interpret, and recall information that aligns with their beliefs [96]; and *cognitive dissonance* leads to the avoidance of information deemed incongruent to one's beliefs [42]. These biases and tendencies can be triggered by the information content that conveys an ideologically polarising issue and, more importantly, can be amplified by algorithmic information curation, which tends to optimise and cater predominantly to the users' preferences and beliefs, even when these may be misinformed [7, 59, 80].

A variety of approaches has been proposed to mitigate cognitive biases using behavioural interventions, such as *nudging* [17] or *boosting* [93]. However, research has suggested that they are not always effective [12, 114]. Specifically, there is no one-size-fits-all solution for debiasing because a person's reaction to debiasing approaches is affected by a variety of individual and contextual factors [2, 49, 89, 110]. Moreover, research has highlighted **individual factors** that dictate why some individuals might be particularly susceptible to exhibiting cognitive biases while others are more resistant. Research in psychology has suggested that individual differences influence how people perform reasoning [127] and how receptive they are to cognitive bias interventions [40, 89]. In the realm of misinformation, studies [33, 104, 134] have shown that individual differences in effortful thinking styles, assessed by the Cognitive Reflection Test (CRT) [44], can predict an individual's susceptibility to misinformation. At the same time, people may react to information stimuli differently, depending on the context and situation of the interaction. Furthermore, there are **contextual factors**, which describe the relationship between the user and the triggers of cognitive biases. For example, an individual's interest and involvement in the topic have been shown to influence how they interact with information [86, 87] and, more importantly, the extent to which they are susceptible to cognitive biases [11, 146]. Moreover, several studies have suggested that attitude strength and attention are essential in activating cognitive biases [1, 109].

By studying susceptibility factors for cognitive biases, we can pave the way for designing more effective bias mitigation techniques that adapt to individuals and interaction contexts. Yet, limited studies have investigated how these factors come into play in human-computer interaction. In this research, we tackle the question – "*How do individual and contextual factors influence the occurrences of cognitive biases* ?" We assess the interplay of individual and contextual factors that influence the manifestation of cognitive biases and the degree to which how people are susceptible to them when interacting with computing systems. To closer study this, we operationalise *confirmation bias*, which presents a tendency of people to rely on their attitudes and ideological beliefs when seeking, interpreting, and recalling information [96, 99, 144]. Confirmation bias is one of the most prominent forms of cognitive bias and is highly prevalent in information consumption [5, 61, 146]. We conducted a user study that exposed participants to information on controversial, divisive issues. We asked them to rank headlines according to reading preference, read and recall a news feed, and evaluate the reliability of individual tweet-like information. Through regression analyses, we examined the interaction effects between confirmation bias, *i.e.*, reliance on prior beliefs when evaluating information, and bias susceptibility factors. Specifically, we investigated which and how individual and contextual factors might *amplify* the effects of confirmation bias in three information consumption scenarios: information-seeking intention, information recall, and information interpretation.

We found that the tendency for effortful thinking, strong political beliefs, and strong issue strength of the content (perceived by the study participants) amplified the effects of confirmation bias (Figure 1). Specifically, ideologically polarised information tended to stand out more in people's memories, especially when it confirms their ideological beliefs. Individuals holding strong political beliefs tended to let their information consumption behaviours be guided by their attitudes. In addition, we found that the design and modality of the task influenced the occurrence of confirmation bias. In summary, this work makes the following contributions:

(1) We present an empirical investigation into individual and contextual factors that make users susceptible to falling for their confirmation bias during information consumption.
(2) We provide a discussion of how interventions can be designed to effectively mitigate the effects of confirmation bias by taking into account the bias susceptibility factors based on the characteristics of the users, the content, and the interaction between them. We also discuss ethical and practical implications for media platforms when incorporating bias susceptibility into intervention designs.

2 Related Work

Our work is grounded in research on behavioural psychology in the context of recent discussions in Human-Computer Interaction (HCI) regarding the interplay of cognitive biases, computing systems, and their users.

2.1 Cognitive Bias

In the 1950s, psychologist Herbert Simon proposed the concept of *bounded rationality* – human rationality is inherently limited [121]. Given the complexity of the world and information present to them, humans apply *mental shortcuts* or *heuristics* to make faster but less deliberate decisions. Amos Tversky and Daniel Kahneman later extended Simon's concept of bounded rationality into the notion of *cognitive bias* [141], where they laid out how mental shortcuts systematically skew humans behaviours from the norm of rational judgment without their awareness. Psychologists and behavioural scientists have documented different forms of cognitive biases. For example, anchoring bias presents a tendency where people rely on the first piece of information they see [141], or availability bias makes individuals rely on information that is mostly available to them [140].

Subsequent research in psychology has augmented the original definition of cognitive biases as *features of the human mind* to cope with the complexity of the world. The prominent psychologist Gerd Gigerenzer viewed that humans apply heuristics as cognitive strategies to effectively make fast decisions [48]. Lieder et al. [88] argued that cognitive biases are mechanisms that humans use to make optimal decisions under their limited cognitive resources. From the lens of evolution psychology, Haselton et al. [55, 56] suggested cognitive biases are inherent mechanisms humans employ as part of their survival and natural adaptation. More importantly, individuals form heuristics or their "rules of thumb" based on the beliefs and preferences they learned from past experiences of the world.

In the realm of online information consumption, users generally have cognitive biases as their inherent, unconscious cognitive strategies for effectively skimming through the sheer volume of information on their news feed and stopping at the news piece of their interest. Different forms of cognitive biases, therefore, affect how humans perceive and evaluate information. For example, confirmation bias [96] and cognitive dissonance [42] prompt individuals to favour information that aligns with their beliefs and avoid what is deemed incompatible. Others, like the continued influence effect [83], make individuals stick to false information although it has been retracted, while negativity bias [75] triggers stronger attentional and emotional responses to information with a negative affect. These cognitive biases become problematic as they enhance the ability of misinformation to deceive people, be disseminated, and persist in memory [6, 18, 117].

Recent discussions in HCI [12, 91] have brought attention to the role of algorithms and recommendation systems in amplifying cognitive biases in users. Different forms of cognitive bias prompt users to seek and expose themselves to information favouring their beliefs. At the same time, recommendation systems optimise on and cater predominantly to the users' preferences and beliefs [9], resulting in amplifying their existing cognitive biases [7, 59, 80]. Without proper intervention, cognitive biases and recommender algorithms together form a self-reinforcing loop, hinder users' ability to make an informed decision, and make them vulnerable to manipulation [3].

A growing body of work has explored how the adverse effects of cognitive biases could be mitigated [12, 110]. Prior research has investigated various debiasing techniques, such as nudging [17, 118, 137] (tapping into people's cognitive biases to shift them towards a desirable behaviour outcome), boosting [93, 111] (nurturing people's metacognitive skills), or decision-support systems [147] (guiding users to make informed, optimal decisions). However, effectively mitigating cognitive biases is challenging, mainly because some individuals are more susceptible to cognitive biases than others [47, 91, 110]. In other words, no one-size-fits-all solution exists: different individuals possess different mental models of interacting with information [2, 49], and, therefore, factors that drive their cognitive biases could be different. Limited research has explored how user- and context-related characteristics come into play with regard to cognitive biases in the context of social media. We review these factors in the following sections, focusing on confirmation bias.

2.2 Confirmation Bias

Confirmation bias presents a tendency to seek, interpret, and recall evidence in a way that they are partial to beliefs, preferences, or hypothesis in hand [96, 99, 148]. It is a long-established phenomenon in psychology [74] and one of the most prominent forms of cognitive biases [38, 98]. In his seminal work, Nickerson [96] demonstrated that confirmation bias occurs largely in humans' everyday decision-making without their awareness, for example, the tendency for people to make a hypothesis about number patterns (*i.e.*, number mysticism [148]), the tendency for doctors to find evidence to support their medical diagnosis, or the tendency for jurors to interpret ambiguous evidence pieces in favour of their pre-existing beliefs.

Notably, confirmation bias overlaps with related phenomena in psychology like motivated reasoning [73, 133]. While both emerge from individuals' reliance on the ideological congruence between their beliefs and the information, each pursues a different scope. Confirmation bias is primarily an unconscious cognitive mechanism that reinforces one's existing beliefs. Meanwhile, motivated reasoning refers to a goal-driven tendency (*e.g.*, to defend one's ideology or values [30]) to favour evidence that confirms one's beliefs while rejecting information deemed unfit. The latter is broader in scope as motivated reasoning can also be deliberately driven by goals and emotions in the reasoning process, as well as subconsciously influenced by confirmation bias [54, 68]. While these phenomena pursue different mechanisms, they both influence how individuals seek, perceive, and recall information. In the realm of information consumption, confirmation bias and motivated reasoning have been attributed to political polarisation [68, 133] and the spread of misinformation [21], as well as giving rise to *selective exposure* [46] (known as individuals' tendency to expose themselves to predominantly information that confirms their beliefs [43]).

In this paper, we investigate factors that influence how people are susceptible to cognitive biases through the lenses of confirmation bias. In conjunction, we operationalise (1) confirmation bias

in the context of information consumption as when users *rely* on the congruence between their beliefs and the ideological stance of the content present to them on a news feed, and (2) user- and interaction-context-related characteristics that influence how cognitive biases manifest.

2.3 The Occurrences of Confirmation Bias in Information Consumption

Confirmation bias exists in many stages of information consumption: it prompts individuals to rely on their attitudes and beliefs, affecting how they seek, perceive, and remember information they encounter. The effects of confirmation bias, therefore, distort individuals' psychological expression (*e.g.*, perception [144], cognitive load [119], and recall [45]), behavioural expression (*e.g.*, clicks [134] and attention [131]), or physiological expression (*e.g.*, peripheral and brain signals [11]). Following Vedejová and Čavojová [144], in this work, we operationalise three scenarios where confirmation bias can manifest: information-seeking intention, interpretation, and memory recall. Next, we briefly review related works that have investigated the effects of confirmation bias in each of the information scenarios.

2.3.1 Information Seeking. Research has studied confirmation bias in information seeking from the lenses of selective exposure [20, 43, 74]. Recent research in HCI has studied the same phenomenon in the online context as users tend to exhibit information selection behaviours in favour of belief-congruent information [86, 106, 112, 134]. For example, Liao and Fu [86] and Pothirattanachaikul et al. [106] showed that people clicked to read more predominantly content items that confirmed their beliefs. On the other hand, Tanaka et al. [134] found that users tended to avoid clicking on fact-checking messages that contradict their pre-existing beliefs.

2.3.2 Information Interpretation. Confirmation bias prompts individuals to evaluate congruent information differently from dissenting information. In psychology, Kobayashi [76] found that individuals tended to put more scrutiny on information against their beliefs. Research in HCI has also studied how users are influenced by their beliefs as a heuristic when evaluating information. van Strien et al. [143] conducted an eye-tracking study and found that individuals' strong attitudes can skew how they evaluate the creditability of information on the web. Allen et al. [4] showed that Twitter Birdwatch users preferably challenge fact-checking content from those with whom they disagree politically. In another example, Wischnewski et al. [149] found that individuals tended to perceive Twitter profiles deemed incongruent to their beliefs as bot accounts.

2.3.3 Information Recall. Based on the schema theory, *schemas*, known as the knowledge structure, is built over time from experiences and memories. A memory schema causes different pieces of information to be remembered differently, thus, resulting in the application of confirmation bias [31, 126]. A seminal study [92] reporting on a car crash experiment suggested that the memory of an event can be distorted by the *perception* of the details during the actual event. A small number of studies, however, have investigated how confirmation bias affects the recall of information and produced mixed results. Some studies have suggested that individuals better recall information that supports their attitudes or beliefs [45, 50, 66, 95]. For example, Frost et al. [45] conducted a study where participants were asked to recognise social media posts. They found that the recognition memory for information congruent with their viewpoints was better than that for dissenting information. Meanwhile, some works have found opposite results. For example, in their first study, Lescarret et al. [81] reported that middle school students tended to recall better attitude-inconsistent information, while there was no such effect in university students. Other research suggested no difference in the recall ability for information supporting or opposing one's prior attitudes [62, 130, 144].

2.4 Influencing Factors for the Occurrence of Cognitive Biases

2.4.1 Individual Factors. Research has pointed out several factors governing individuals' tendency to fall victim to cognitive biases. One of the most prominent indicators is the individual difference in *effortful thinking styles*. Cognitive biases are byproducts of using our intuitive, fast System 1 thinking instead of the deliberate but slower System 2 thinking [39, 69]. Research has shown that individual differences in effortful thinking styles correlate with the occurrences of cognitive biases [128, 136, 145] and the discernment of misinformation [8, 85, 104]. Some of the effortful thinking indicators include the Cognitive Reflection Test (CRT) [44], which measures one's tendencies to use intuitive thinking over deliberative thinking, the Need for Cognition Scale (NFC) [15], which gauges the tendency to engage in effortful cognitive activities, and the Bullshit Receptivity Scale (BRS) [102], which reflects the ability to detect *bullshit* or statements with profound meaning. Recent research in HCI has increasingly used these techniques to investigate the relationship between such factors and how people interact with information online [70, 134].

Moreover, political attitudes also affect individuals' receptivity to cognitive bias interventions. Research in psychology has shown that individual differences in political ideology influence how they process information [36, 120]. In addition, studies have shown that people who leaned towards conservative beliefs were more likely to exhibit less reflective thinking [25] and be more resistant to misinformation correction than individuals on the liberal end of the political spectrum [34, 35, 51]. Yet, recent research has argued that this tendency did not hold exclusively for political conservatives, as inclinations for liberal beliefs [37] or any of the political extremes [142] render people susceptible to conspiracy thinking.

2.4.2 Contextual Factors. The occurrences of cognitive biases also depend on the contextual relationship between the user and the information. In one direction, an individual's interest or involvement with the information describes such relationships. This has been highlighted as a moderating or amplifying factor for cognitive biases in prior HCI studies [11, 86, 87, 146]. In behavioural psychology, Richter et al. [108, 109] proposed the Two-Step Model of Validation, which explains that individuals with relevant background knowledge tended to process conflicting information more elaboratively. In other words, they may bypass the use of cognitive biases towards a more balanced way of information processing.

The information's ability to amplify or trigger cognitive biases explains the other end of this contextual relationship as well. Research has pointed out its ability – *e.g.*, strong language and slant – to trigger people's negative emotional responses [94] and attention [1, 115], thus, making them susceptible to believing and sharing false information. Recent works in HCI have also highlighted linguistic and sentimental features that prompt strong emotional responses and potentially trigger people's mental shortcuts [123, 151].

2.5 Our Contributions

People adopt cognitive biases to effectively sift through a large amount of information presented to them when they roam through online platforms. This comes at a cost; cognitive biases prompt their irrationality and make them vulnerable to manipulation. While it is important to mitigate the adverse effects of cognitive biases, recent research in HCI has suggested that bias mitigation is challenging because cognitive biases manifest differently according to various innate user characteristics and contexts of the interaction between users, information, and systems. Our work contributes to HCI research as we investigate the influence of individual and contextual factors on cognitive bias susceptibility. We consider three relevant scenarios of information consumption where confirmation bias manifests: (1) information-seeking intention, (2) interpretation, and (3) recall. Accordingly, we employ three representative tasks for information consumption: (1) headline ranking, (2) individual tweet evaluation, and (3) news feed-free recall. Based on our findings, we shed light on implications for designers and media platforms tailoring effective interventions that consider bias susceptibility in users, information, and their interaction.

Additionally, this work contributes new knowledge to the literature by looking at the influence of bias susceptibility factors on recall ability, on which there has been limited research. In addition, to the best of our knowledge, our work is the first to assess the effects of confirmation bias on memory recall using a *delayed free recall* task, where participants are presented with a sequence of tweet-like information and subsequently, after some delays, asked to recall them in any order.

3 Methodology

We conducted a user study to explore indicators for susceptibility to confirmation bias. Therefore, we exposed participants to a number of tweets stating opinions on controversial topics. Subsequently, they engaged in three tasks: ranking headlines, recalling tweets on a news feed, and evaluating tweets.

3.1 Study Stimuli

During the experiment, we showed tweets containing only textual information (*i.e.*, no images) on three controversial, polarising topics. Each tweet aligned with one end of the ideological spectrum, *i.e.*, supporting (pro) or opposing (con). We picked tweets concerning the following topics: *abortion rights*, *same-sex marriage*, and *vegetarianism*. All chosen topics have been widely debated globally with increasingly polarised viewpoints [23, 57, 101] and lend themselves, therefore, well to our study.

We sourced all tweets from the ProCon.org website[1], which provides facts, opinions, and arguments on various controversial topics on both ends of the ideological spectrum. For example, on the abortion rights issue, the *pro* stance endorses the idea that *abortion should be legal*. In contrast, the *con* stance supports the idea that *abortion should be prohibited*. Table 2 shows pro-con ideology pairs for each topic deployed in this study. In addition, we made sure all tweets were in English and between 40 to 70 words in length to resemble the standard 250-character tweets. Following the approach in related works [11, 29, 112], we hypothesise that the tweets would trigger participants' confirmation biases by making them rely on their pre-existing beliefs when assessing the information.

We gathered eight tweets on each of the three topics, consisting of four pro and four con tweets. Each tweet was presented on the screen with the same font, compact line spacing, alignment, column width, colour text (black), and white background (see Figure 2, middle item). We deployed our stimuli and questionnaire on Qualtrics[2]. We did not provide source information in our stimuli to separate confounds such as source bias [135].

3.2 Study Design

3.2.1 Experimental Design. To study the effects of contextual and individual factors on the occurrences of confirmation bias, we conducted a study with a within-subject design. We measured the effects of confirmation bias (in three scenarios: information-seeking intention, information recall ability, and information interpretation ratings) and investigated the influence of the following predictors: the ideological congruence score between the user and tweet, individual factors (as measured by the Cognitive Reflection Test (CRT), Need For Cognition scores (NFC), Bullshit Receptivity Scales (BRS), and Wilson-Patterson Conservatism Scales (WPCS)), and contextual factors (topic interest, and tweet perceived issue strength). Table 1 summarises all predictor and measurement variables we examined in this study. We determined the required sample size ($N = 42$) by a priori power analysis using *G*Power* [41] with a medium-to-large effect size $f^2 = 0.25$, power $1 - \beta = 0.80$, type I error probability $\alpha = 0.05$, and two predictors.

3.2.2 Participants. We invited 42 participants (16 men, 25 women, and one non-binary) through the university network to join the study in our usability lab. All participants were native or fluent speakers of the English language, and their mean age was 28.51 years ($SD = 8.52$), with the minimum and maximum ages being 19 and 54 years old, respectively. Of 42 participants, 11 held a postgraduate degree, 19 held a bachelor's degree, and the remaining 12 participants had at least 12 years of education.

3.2.3 Procedure. The study took place in a quiet room in our institution's usability lab. We first informed each participant about the objective and procedure of the study and collected their written consent. We seated participants before a screen and asked them to adjust their seating to a comfortable position. Participants responded to a pre-study survey collecting information on individual

[1] www.procon.org
[2] www.qualtrics.com

Table 1: List of the examined predictor and measurement variables.

Variable	Measure	Scale
Predictor Variables	**Ideological Congruence**	
	- Implicit Congruence (Cong_Imp) [77, 86]	Ordinal (-7 to +7)
	Individual Factors	
	- Cognitive Reflection Test score (CRT) [44]	Count of correct responses (0 to 7)
	- Need For Cognition scale (NFC) [15]	Ordinal (5-Likert scale: 1 to 5)
	- Bullshit Receptivity Scale (BRS) [102]	Ordinal (5-Likert scale: 1 to 5)
	- Wilson-Patterson Conservatism Scale (WPCS) [58]	Count of conservative items (−27 to +27)
	Contextual Factors	
	- Topic Interest (Interest) [63, 86]	Ordinal (7-Likert scale: 1 to 7)
	- Stimulus Perceived Issue Strength (Strength)	Ordinal (5-Likert scale: 0.5 to 4.5)
Measurement Variables	**Information-Seeking Intention**	
	- Headline Rank Position	Ordinal (1 to 8)
	Information Interpretation	
	- Information Interpretation scale [73, 144]	Ordinal (5-Likert scale: 1 to 5)
	Information Recall	
	- Recall Ability Score	Ordinal (4-Likert scale: 0 to 3)

Table 2: Topics presented and their ideological ends.

Topic	*Pro* stance	*Con* stance
Vegetarianism	People should become Vegetarian	People should not become Vegetarian
Abortion Rights	Abortion should be legal	Abortion should be prohibited
Same-sex Marriage	Same-sex marriage should be legal	Same-sex marriage should be prohibited

and contextual factors with their demographic information (see Section 3.3.1), after which they were asked to complete the following tasks:

(1) **Ranking Headlines**: Participants were presented with a list of tweet headlines (eight per topic), each of which was a one-sentence snippet of a tweet. The headlines were initially presented in a random order. The participants had to reorder the headlines from what they wanted to read the most (top, first place) to what they wanted to read the least (bottom, last place).
(2) **Reading and Recalling Tweets**: Subsequently, participants were presented with a news feed of eight tweets in a random order. They could scroll up and down to read each tweet. Once they finished reading the tweets, participants were asked to engage in an interruption task where they had to solve seven summations (adding two single-digit numbers). Inspired by previous studies that employed recall tasks [62, 125], the interruption task was introduced to separate participants mentally from the tweets and to reset their working memory, as well as to serve as attention checks. After that, participants responded to a free-recall task, namely "*please write down every viewpoint, aspect, and detail that you can remember from the tweets you have just read.*" Participants were given a maximum of five minutes to type in their responses using a keyboard. They were told to ignore spelling and grammatical errors.
(3) **Evaluating Individual Tweets**: Lastly, participants were again presented with the earlier shown tweets and asked to rate the tweet according to the information interpretation scales and how polarised it was (See Section 3.3.2). Each tweet was presented with the in-study survey on the same page. Once they finished the survey, participants could click "Next" and proceed to the subsequent tweet.

We repeated all three tasks for each topic (abortion rights, same-sex marriage, and vegetarianism) and incorporated eight tweets (four pros and four cons) per topic. The presentation order of topics was randomised, while the tasks were repeated in a fixed order. A short break between each topic allowed participants to relax for at least 15 seconds before proceeding to the next topic. Figure 2 visualises the procedure of our study.

Upon completion, we compensated each participant with a $20 electronic voucher for their time. The study took about an hour to complete and was approved by the Human Research Ethics Committee of the University of Melbourne.

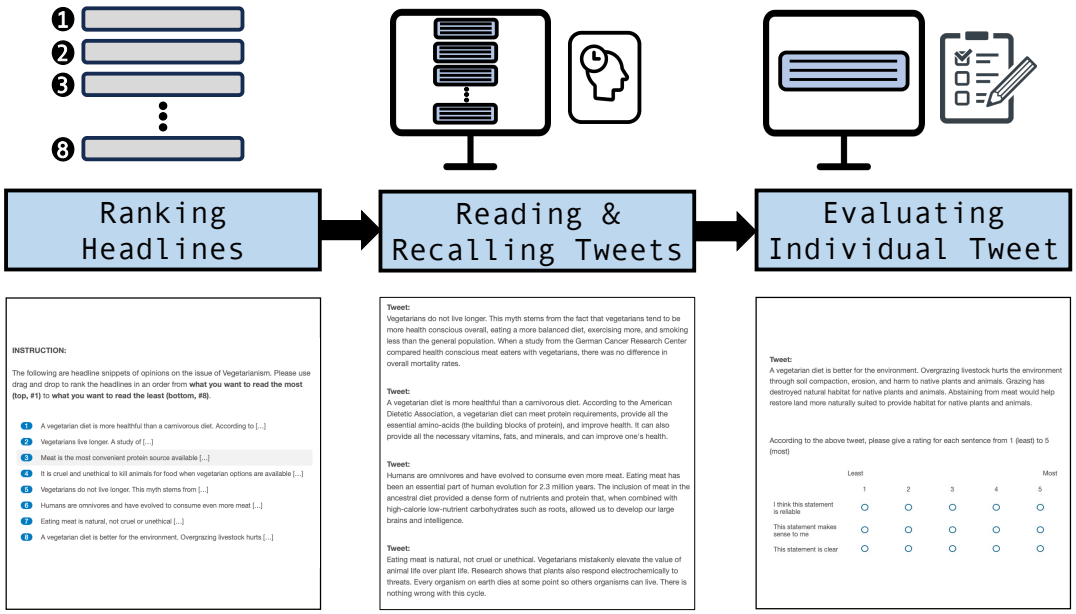


Figure 2: Summary of the study procedure. For each topic, participants completed three tasks in the following order: Headline Ranking, Reading and Recalling Tweets, and Individual Tweet Evaluation. Example screenshots of the tasks are shown below.

3.3 Surveys

3.3.1 Pre-study Survey. We gauged potential individual and contextual factors for cognitive biases through the pre-study survey, which was administered after the participant provided informed consent. The survey started with four tests: **Cognitive Reflection Test (CRT)**, **Need For Cognition scores (NFC)**, **Bullshit Receptivity Scales (BRS)**, and **Wilson-Patterson Conservatism Scales (WPCS)**. We derived the 7-item version of CRT [122] consisting of seven binary-choice questions gauging the participant's tendency to use System 1 over System 2 thinking. For each question, there was one correct answer and one wrong answer. NFC consists of questions asking participants to self-report their tendency to use System 1 thinking on a 7-Likert Scale (1: extremely uncharacteristic to 7: extremely characteristic) [15]. For the BRS, we followed the approach in Pennycook et al. [102], presenting participants with a list of 10 profound statements in English, asking them to rate how profound they thought each statement was on a 5-Likert scale (1: least profound and 5: most profound). Finally, for the WPCS, we presented participants with 27 socio-economic policies and asked them to indicate whether they agreed, disagreed, or felt neutral with each policy. We opted for the scales by Henningham [58], developed exclusively for our study population in Australia. The number of liberal and conservative policies were counterbalanced, and their presentation order was randomised. For each of the four tests, the presentation order was randomised.

In addition, for each of the three topics, we asked participants to rate how they felt about the following statements on a 7-Likert scale (1: strongly disagree to 7: strongly agree): "*this issue is related to my core values*," "*it is important to defend my point of view on this issue*," "*I am interested in learning about this issue*," and "*I desire to know the facts about this issue*." Following the approach in [63, 86], we took the average of their four responses to measure the participants' **topic interest**. We then measured participants' ideological stance on the topic in a self-report question and a word association test. In the self-report question, we asked participants to rate their stance on a continuous scale from 0 to 100 (0: strongly opposing to 100: strongly supporting). In the word association test, participants were presented with each topic name (*i.e.*, abortion rights, same-sex marriage, and vegetarianism) and asked to associate each of them with five pairs of bipolar adjectives: *unfavourable-favourable*, *bad-good*, *unnecessary-necessary*, *harmful-beneficial*, and *unhealthy-healthy*. Following the same test deployed in prior studies [77, 86], we asked participants to choose their stance on a 7-Likert scale (1: negative adjective to 7: positive adjective).

3.3.2 In-study Survey. For each tweet presented, we gauged how participants interpreted the tweet using the information interpretation scales derived from Klaczynski [73]. We asked them to rate the following statements on a 5-Likert scale (1: strongly disagree to 5: strongly agree): "*This tweet is reliable to me*", "*This tweet is clear to me*", and "*This tweet makes sense to me*". In addition, we asked participants to rate how polarised the tweet was according to its expressed ideology. On a 10-Likert scale, we asked them to rate whether the tweet supports the pro or con stances (1: strongly supporting the con stance to 10: strongly supporting the pro stance).

3.4 Predictor Variables

3.4.1 Individual Factors. We derived the following individual factors: CRT, NFC, BRS, and WPCS. We calculated the CRT scores by counting the number of correct responses. Higher CRT, therefore, indicates a higher tendency for effortful thinking [44]. For NFC and BRS, we averaged the internal question items, 6 for NFC and 10

Table 3: Descriptive statistics (Mean, S.D., max, min) of the topic-wise ideological stances and contextual factors collected in our study

Factors	*Abortion Rights*	*Vegetarianism*	*Same-sex Marriage*
Stance_Imp (7-Likert scale: [1..7])	M = 5.13, S.D. = 1.65, max = 7, min = 1	M = 3.42, S.D. = 1.64, max = 6.25, min = 1	M = 4.40, S.D. = 1.96, max = 7, min = 1
Stance_Exp (continuous: [0..100])	M = 86.76, S.D. = 22.21, max = 100, min = 1	M = 52.05, S.D. = 21.17, max = 100, min = 0	M = 79.92, S.D. = 32.79, max = 100, min = 1
Interest (7-Likert scale: [1..7])	M = 5.14, S.D. = 1.65, max = 7, min = 1	M = 3.42, S.D. = 1.65, max = 6.25, min = 1	M = 4.40, S.D. = 1.96, max = 7, min = 1
Strength (overall) (5-Likert scale: [0.5..4.5])	M = 3.46, S.D. = 1.22, max = 4.5, min = 0.5	M = 3.20, S.D. = 1.31, max = 4.5, min = 0.5	M = 3.29, S.D. = 1.38, max = 4.5, min = 0.5
Strength (pro-tweet only)	M = 3.46, S.D. = 1.23 max = 4.5, min = 0.5	M = 3.50, S.D. = 1.21 max = 4.5, min = 0.5	M = 2.87, S.D. = 1.43 max = 4.5, min = 0.5
Strength (con-tweet only)	M = 3.47, S.D. = 1.22 max = 4.5, min = 0.5	M = 2.92, S.D. = 1.35 max = 4.5, min = 0.5	M = 3.73, S.D. = 1.18 max = 4.5, min = 0.5

for BRS. Cronbach's alpha of 0.731 and 0.819 showed acceptable and good consistency for NFC and BRS, respectively. Lastly, we obtained WPCS from the sum of the agreeing responses to each liberal policy and the disagreeing responses to each conservative response. Thus, positive WPCS indicates a higher inclination towards liberal ideologies, while negative WPCS indicates a higher inclination towards conservative ideologies. The distributions of each measure were as follows: CRT (M = 4.21, S.D. = 1.97, max = 7, min = 0), NFC (M = 3.56, S.D. = 0.42, max = 4.33, min = 2.50), BRS (M = 3.17, S.D. = 0.77, max = 4.60, min = 1), and WPCS (M = 9.33, S.D. = 6.32, max = 21, min = −5). We found that the distribution of WPCS was skewed towards strong liberal ideologies. The statistics suggest that 16 individuals scored above 10, 24 scored between 1 and 10, and 2 scored between -5 and 0. This implies that most of our participants held moderate liberal ideologies while the lower end of our population implied those who held either neutral or leaned slightly towards conservative ideologies.

3.4.2 Contextual Factors. We obtained the participants' topic interest levels, Interest, which showed good internal consistency among the 4-item questions (Cronbach's alpha = 0.87). In addition, we obtained the perceived issue strength score, Strength, of each tweet evaluated by the participant by taking the unsigned distance between the self-report tweet issue strength (TP: 10-Likert scale, from 1: strongly opposing to 10: strongly supporting) and its neutral absolute (score of 5.5), *i.e.*, Strength = |TP − 5.5|. Table 3 summarises the statistics of the topic interest levels and the stimulus perceived issue strength score for each topic reported by our participants.

3.4.3 Ideological Stance and Congruence. We derived the participant's ideological stance in two ways: **implicit** (Stance_Imp) collected from the word association test (7-Likert scale), and **explicit** (Stance_Exp) collected from a self-report question (continuous from 0 to 100). The statistics of the implicit and explicit stances are also shown in Table 3. We found that most of our participants rather expressed *pro* attitudes for abortion rights and same-sex marriage while showing neutral stances on the issue of vegetarianism.

Subsequently, we derived the ideological congruence score of the user stimulus from the product of the ideological stances of the user and the tweet. Denoting $\textsc{Stance}(T)$ as the ideological stance of tweet T (-1: supporting the *con* stance and +1: supporting the *pro* stance), we calculated the explicit congruence (Cong_Exp) and implicit congruence (Cong_Imp) between participant P and tweet T using the following equations.

$$\textsc{Cong\_Imp}(P,T) = \textsc{Stance\_Imp}(P) \times \textsc{Stance}(T)$$
$$\textsc{Cong\_Exp}(P,T) = (\textsc{Stance\_Exp}(P) - 50) \times \textsc{Stance}(T)$$

We cross-checked Cong_Imp with Cong_Exp and found a Pearson correlation of 0.946, indicating that they were highly correlated. This ensures our internal validity as participants' self-assessments aligned with the implicit measures. Therefore, we report our analysis using Cong_Imp, *i.e.*, the implicit measure for ideological congruence.

3.5 Measures

3.5.1 Information-Seeking Intention. We derived the order of each tweet headline directly from the participant's final ranking of eight headlines. Each headline was labelled between 1 and 8, where 1 and 8 represented the most desirable headline to read and the least desirable headline to read accordingly.

3.5.2 Information Recall. We assessed the recall ability from the written, recalled responses in the free-recall task. In particular, we rated how well the response matched the presented tweets. Two researchers independently coded each recalled response according to the content of each tweet mentioned. Subsequently, for each matched tweet, they individually gave a rating for the richness of the response on a 4-item Likert-style scale (0: little or no mention of the tweet, 1: somewhat rich, 2: moderately rich, and 3: very rich). In particular, we scored 3 for the recalled tweet if it stated all aspects and details and closely resembled the original tweet.

A score of 2 was given if the recalled tweet was incomplete but featured more than one aspect or detail. A score of 1 was given if the recall mentioned only one aspect of the original tweet. A score of 0 was given if the tweet was not mentioned in the recalled response. We achieved an inter-rater reliability (Cohen's Kappa) of 0.779, indicating good consistency between the two raters. In summary, for each tweet, we derived a measure for the recall ability as a floor average of the ratings from two raters, *i.e.*, $\texttt{Rating}(T) = \lfloor(\texttt{Rating}_{R1}(T) + \texttt{Rating}_{R2}(T))/2\rfloor$.

3.5.3 Information Interpretation. Among the three information interpretation rating items (5-Likert scale) collected in the individual tweet evaluation task, a Cronbach's alpha of 0.783 indicated acceptable internal consistency across the three items. Therefore, we used an average of the three items as our measure.

4 Results

Given that the derived measures are all ordinal data, we performed mixed-effect ordinal regression analyses using the Cumulative Linked Mixed Models (CLMM) [19] to assess the effect of the user-stimulus ideological congruence and individual and contextual factors. Furthermore, we examined the interplay of individual and contextual factors with the reliance on ideological congruence as a *heuristic* for confirmation bias. To do so, we assessed the interaction effects between the ideological congruence, Cong_Imp and one of the individual or contextual factors in three scenarios that aimed at stimulating confirmation bias [144]: headline ranking, news feed free-recall, and individual tweet evaluation.

We performed the regression analyses to examine the main and interaction effects between Cong_Imp and each of the predictors: (a) Interest, (b) CRT, (c) WPCS, (d) Strength, (e) NFC, and (f) BRS. To avoid colinearity issues, we ran the regression analysis separately for each measure and predictor. We also accounted for random effects from participants and stimuli to reflect our repeated-measure study design. Therefore, the formula for our CLMM models is `(measure ~ 1 + cong_imp * predictor + (1|participant_id) + (1|stimulus_id))`. We standardised all predictors before performing the regression analyses. From the data collected, we discarded 29 observations (2.87%) due to data losses (*e.g.*, the Qualtrics survey did not successfully capture some of the participants' responses). We note that our statistical models (CLMMs) are robust against missing data, which was minimal in our experiment.

Further, we looked into each interaction effect by performing posthoc regression analyses using CLMM to compare the estimated regression coefficient (β) or the main effect size of ideological congruence (Cong_Imp) between two different conditions of the interaction predictor variables, separated by its median value. Following the approach in Clogg et al. [22], we then performed a two-sampled Z-test to compare the regression coefficients (*i.e.*, the main effect size of Cong_Imp on the measurement variable) between two conditions (High and Low). The formula for the Z statistic is denoted in Equation 4, where β, σ, and n represent the estimated regression coefficient (main effect size), the standard error, and the sample size, respectively.

$$Z = \frac{\beta_{\text{High}} - \beta_{\text{Low}}}{\sqrt{\sigma^2_{\text{High}}/n_{\text{High}} + \sigma^2_{\text{Low}}/n_{\text{Low}}}}$$

In this section, we present the results of the regression and posthoc analyses, which reveal the relationships between the occurrences of confirmation bias and their influencing factors. Tables 4, 5, and 6 show a brief summary of regression results for information-seeking intention, recall, and interpretation, respectively. Table 7 depicts test statistics from the posthoc regression analyses and coefficient comparisons. We include the full regression tables in the supplementary materials.

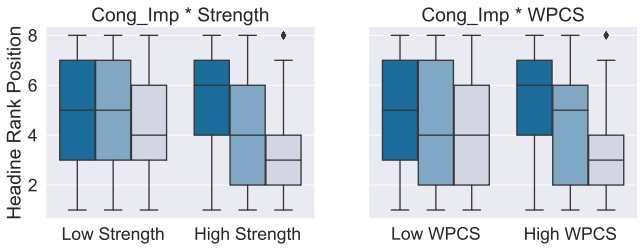
4.1 Information-Seeking Intention

4.1.1 Regression Analyses. We ran a mixed-effect ordinal regression with the ranking position of each tweet headline as the dependent variable on the data collected from the headline ranking task (Models 1a-1f). We found significant interaction effects of Cong_Imp × WPCS ($p < 0.001$, $\beta = 3.841$, $S.E. = 0.801$, $Z = 4.798$) and Cong_Imp × Strength ($p < 0.001$, $\beta = 1.334$, $S.E. = 0.581$, $Z = 2.295$). We did not detect significant interaction effects of CRT, NFC, BRS, and Interest with the implicit ideological congruence. Table 4 summarises the regression results for headline rank position. Figure 3a displays boxplots that illustrate the interaction effects of Cong_Imp × Strength and Cong_Imp × WPCS with x-axes being the interaction variables in two conditions, high (above or equal median) and low (below median).

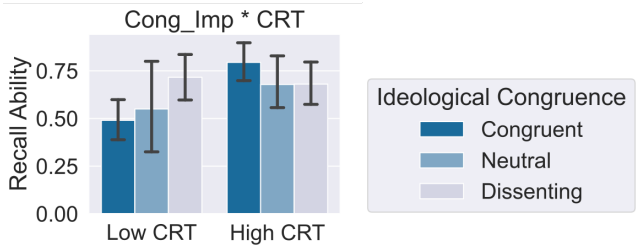
4.1.2 Posthoc Analyses. To closely examine the interaction effects, we performed posthoc comparisons of regression coefficients. On the interaction effect Cong_Imp × WPCS, we found that participants who held a relatively strong liberal ideology tended to rely more on ideological congruence than those more moderately oriented ($p < 0.001$, $Z = 52.076$; High WPCS: $p < 0.001$, $\beta = 2.618$, $S.E. = 0.431$, $Z = 6.073$; Low WPCS: $p < 0.001$, $\beta = 1.183$, $S.E. = 0.436$, $Z = 2.711$). Subsequently, posthoc analyses on the interaction effect Cong_Imp × Strength suggested that participants relied on ideological congruence for headlines of ideologically strong tweets more than those of ideologically neutral tweets ($p < 0.001$, $Z = 40.968$; High Strength: $p < 0.001$, $\beta = 2.305$, $S.E. = 0.408$, $Z = 5.698$; Low Strength: $p < 0.001$, $\beta = 1.196$, $S.E. = 0.408$, $Z = 2.931$).

4.2 Information Recall

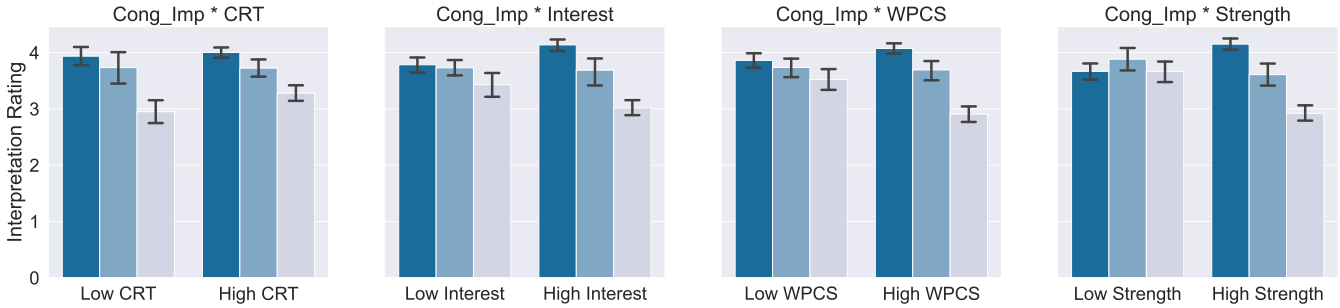
4.2.1 Regression Analyses. We performed a mixed-effect ordinal regression on the recall ability scores (Models 2a-2f). We only detected significant interaction effects of Cong_Imp × CRT ($p = 0.005$, $\beta = 0.018$, $S.E. = 0.006$, $Z = 2.781$). We did not detect significant interaction effects of NFC, BRS, Interest, and Strength with the implicit ideological congruence. Interestingly, we found a significant main effect of the tweet's issue strength on the recall ability ($p = 0.011$, $\beta = 0.628$, $S.E. = 0.245$, $Z = 2.560$). Specifically, we found individuals tended to recall the tweet better if they perceived it as ideologically stronger. Table 5 summarises the regression results for recall ability. Figure 3b illustrates the interaction effect of Cong_Imp × CRT.



(a) Interaction effects on the headline rank position. Posthoc regression analyses show that the effect of Cong_Imp on the headline rank position in High Strength and High WPCS conditions are stronger than those in Low Strength and Low WPCS, respectively.



(b) Interaction effects on the recall ability. Posthoc regression analyses show that the effect of Cong_Imp on the recall ability is significant in the Low CRT condition, but insignificant in the High CRT condition.



(c) Interaction effects on the information interpretation ratings. Posthoc regression analyses show that the effect of Cong_Imp on information interpretation ratings in High Interest, High WPCS, and Low CRT conditions are stronger than those in Low Interest, Low WPCS, and High CRT, respectively. The effect of Cong_Imp is significant in the High Strength condition, but insignificant in the Low Strength condition.

Figure 3: Interaction effects on (a) the headline rank position (Cong_Imp×Strength and Cong_Imp×WPCS), (b) the recall ability (Cong_Imp × CRT), and (c) the information interpretation ratings (Cong_Imp × CRT,Cong_Imp × Interest,Cong_Imp × WPCS, and Cong_Imp × Strength). In each interaction plot, different levels of Cong_Imp are compared: congruent ([3..7]), neutral ([-2..+2]), and dissenting ([-7..-3]).

4.2.2 Posthoc Analyses. From posthoc comparisons of regression coefficients, we found that individuals who scored lower on CRT tended to show better recall of ideologically dissenting tweets than congruent ones (Low CRT: $p = 0.022$, $\beta = -0.8073$, $S.E. = 0.354$, $Z = -2.280$). For individuals with higher CRT scores, however, we detected no significant linear relationship between the recall ability and the implicit ideological congruence (High CRT: *n.s.*, $\beta = 0.367$, $S.E. = 0.381$, $Z = 0.965$).

4.3 Information Interpretation

4.3.1 Regression Analyses. We performed mixed effect ordinal regression analyses on the information interpretation ratings (models 3a-3f). We found significant interaction effects of Cong_Imp × CRT ($p = 0.003$, $\beta = -1.745$, $S.E. = 0.5911$, $Z = -2.953$), Cong_Imp × WPCS ($p < 0.001$, $\beta = 5.581$, $S.E. = 0.854$, $Z = 6.529$), Cong_Imp × Interest ($p < 0.001$, $\beta = 2.4273$, $S.E. = 0.730$, $Z = 3.324$) and Cong_Imp × Strength ($p < 0.001$, $\beta = 4.930$, $S.E. = 0.646$, $Z = 7.632$). We found no significant interaction effect of NFC and BRS with the implicit ideological congruence. Table 6 summarises the regression results for information interpretation. Figure 3c shows barplots visualising the interaction effects of Cong_Imp × CRT, Cong_Imp × Interest, Cong_Imp × WPCS and Cong_Imp × Strength.

4.3.2 Posthoc Analyses. Posthoc comparisons revealed that, when interpreting information, participants who held a rather extreme ideology tended to rely on ideological congruence more than those with somewhat nuanced ideology ($p < 0.001, Z = 63.858$; High WPCS: $p < 0.001$, $\beta = 3.778$, $S.E. = 0.457$, $Z = 8.259$; Low WPCS: $p < 0.001$, $\beta = 1.910$, $S.E. = 0.524$, $Z = 3.645$). Individuals who exhibited higher topic interest also relied on ideological congruence more than those with lower interest levels ($p < 0.001$, $Z = 65.942$; High Interest: $p < 0.001$, $\beta = 3.302$, $S.E. = 0.439$, $Z = 7.517$; Low Interest: $p < 0.001$, $\beta = 1.450$, $S.E. = 0.620$, $Z = 2.337$). We found individuals who scored lower on CRT tended to exhibit a stronger effect size of ideological congruence than those who scored higher ($p < 0.001$, $Z = -4.055$; High CRT: $p < 0.001$, $\beta = 2.833$, $S.E. = 0.442$, $Z = 6.399$; Low CRT: $p < 0.001$, $\beta = 2.955$, $S.E. = 0.514$, $Z = 5.750$). However, for the issue strength, we only detected a significant effect size of Cong_Imp on information interpretation the high Strength group (High Strength: $p < 0.001$, $\beta = 3.894$, $S.E. = 0.408$, $Z = 9.533$), while the low Strength group showed no significant linear relationship (Low Strength: *n.s.*, $\beta = 0.915$, $S.E. = 0.547$, $Z = 1.673$). In other words, we found that individuals tended to rely on ideological congruence when interpreting tweets with a strong ideological stance.

Table 4: Summary of main and interaction effects between ideological congruence and bias susceptibility factors from the ordinal mixed-effect regression analysis on headline ranking (DV: Rank Position). ** marks a significant effect.

DV: Rank Position	Coef. (β)	S.E.	Z	p-value
Cong_Imp	1.097	0.662	1.656	0.097
Interest	−0.486	0.364	−1.335	0.182
Cong_Imp × Interest	1.110	0.637	1.743	0.081
Model Summary (1a)	Log-Likelihood = −1979.60, AIC = 3983.19, $\text{Cond}_H = 1.9 \times 10^3$			
Cong_Imp	2.438	0.427	5.702	$< 0.001^{**}$
CRT	0.356	0.343	1.040	0.299
Cong_Imp × CRT	−0.672	0.558	−1.202	0.229
Model Summary (1b)	Log-Likelihood = −1980.48, AIC = 3984.97, $\text{Cond}_H = 9.6 \times 10^2$			
Cong_Imp	1.038	0.549	1.889	0.059
Strength	−0.671	0.370	−1.814	0.070
Cong_Imp × Strength	1.334	0.581	2.295	0.022^{**}
Model Summary (1c)	Log-Likelihood = −1920.42, AIC = 3864.84, $\text{Cond}_H = 1.5 \times 10^3$			
Cong_Imp	−0.652	0.665	−0.980	0.327
WPCS	−1.822	0.469	−3.887	$< 0.001^{**}$
Cong_Imp × WPCS	3.841	0.801	4.798	$< 0.001^{**}$
Model Summary (1d)	Log-Likelihood = −1969.39, AIC = 3962.78, $\text{Cond}_H = 2.2 \times 10^3$			
Cong_Imp	1.499	0.591	2.535	0.011^{**}
NFC	−0.447	0.449	−0.995	0.320
Cong_Imp × NFC	0.893	0.738	1.210	0.226
Model Summary (1e)	Log-Likelihood = −1980.48, AIC = 3984.96, $\text{Cond}_H = 1.9 \times 10^3$			
Cong_Imp	2.332	0.505	4.614	$< 0.001^{**}$
BRS	0.138	0.442	0.314	0.753
Cong_Imp × BRS	−0.425	0.714	−0.595	0.552
Model Summary (1f)	Log-Likelihood = −1981.00, AIC = 3985.99, $\text{Cond}_H = 1.5 \times 10^3$			

Table 5: Summary of main and interaction effects between ideological congruence and bias susceptibility factors from the ordinal mixed-effect regression analysis on recall ability scores (DV: Recall Ability). ** marks a significant effect.

DV: Recall Ability	Coef. (β)	S.E.	Z	p-value
Cong_Imp	0.007	0.079	0.094	0.925
Interest	0.018	0.027	0.684	0.494
Cong_Imp × Interest	−0.001	0.006	−0.203	0.839
Model Summary (2a)	Log-Likelihood = −1012.46, AIC = 2040.93, $\text{Cond}_H = 9.7 \times 10^4$			
Cong_Imp	−0.076	0.034	−2.287	0.022^{**}
CRT	0.073	0.069	1.045	0.296
Cong_Imp × CRT	0.018	0.006	2.781	0.005^{**}
Model Summary (2b)	Log-Likelihood = −1008.30, AIC = 2032.60, $\text{Cond}_H = 1.9 \times 10^4$			
Cong_Imp	−0.050	0.045	−1.121	0.262
Strength	0.628	0.245	2.560	0.011^{**}
Cong_Imp × Strength	0.054	0.049	1.111	0.266
Model Summary (2c)	Log-Likelihood = −1008.59, AIC = 2033.19, $\text{Cond}_H = 7.7 \times 10^2$			
Cong_Imp	−0.018	0.040	−0.459	0.647
WPCS	0.035	0.021	1.675	0.093
Cong_Imp × WPCS	8×10^{-4}	0.003	0.319	0.750
Model Summary (2d)	Log-Likelihood = −1011.31, AIC = 2038.62, $\text{Cond}_H = 1.1 \times 10^5$			
Cong_Imp	−0.009	0.124	−0.078	0.938
NFC	0.028	0.055	0.513	0.608
Cong_Imp × NFC	7.638×10^{-5}	0.006	0.014	0.989
Model Summary (2e)	Log-Likelihood = −1012.59, AIC = 2041.17, $\text{Cond}_H = 4.3 \times 10^6$			
Cong_Imp	0.047	0.056	0.845	0.398
BRS	0.013	0.018	0.765	0.444
Cong_Imp × BRS	−0.001	0.001	−1.080	0.280
Model Summary (2f)	Log-Likelihood = −1011.84, AIC = 2039.69, $\text{Cond}_H = 2.4 \times 10^6$			

Table 6: Summary of main and interaction effects between ideological congruence and bias susceptibility factors from the ordinal mixed-effect regression analysis on individual tweet interpretation (DV: Info. Interpretation). ∗∗ marks a significant effect.

DV: Info. Interpretation	Coef. (β)	S.E.	Z	p-value
Cong_Imp	0.875	0.775	1.130	0.259
Interest	−1.404	0.591	−2.376	0.018^{**}
Cong_Imp × Interest	2.427	0.730	3.324	$< 0.001^{**}$
Model Summary (3a)	Log-Likelihood = −2098.41, AIC = 4230.82, $\text{Cond}_H = 4.2 \times 10^3$			
Cong_Imp	4.106	0.498	8.244	$< 0.001^{**}$
CRT	1.317	0.593	2.223	0.026^{**}
Cong_Imp × CRT	−1.745	0.591	−2.953	0.003^{**}
Model Summary (3b)	Log-Likelihood = −2099.26, AIC = 4232.52, $\text{Cond}_H = 2.8 \times 10^3$			
Cong_Imp	−0.529	0.599	−0.885	0.376
Strength	−2.104	0.410	−5.124	$< 0.001^{**}$
Cong_Imp × Strength	4.930	0.646	7.632	$< 0.001^{**}$
Model Summary (3c)	Log-Likelihood = −2072.17, AIC = 4178.35, $\text{Cond}_H = 2.5 \times 10^3$			
Cong_Imp	−0.857	0.714	−1.201	0.229
WPCS	−2.756	0.757	−3.642	$< 0.001^{**}$
Cong_Imp × WPCS	5.581	0.854	6.529	$< 0.001^{**}$
Model Summary (3d)	Log-Likelihood = −2082.53, AIC = 4199.05, $\text{Cond}_H = 4.5 \times 10^3$			
Cong_Imp	2.745	0.676	4.062	$< 0.001^{**}$
NFC	−0.357	0.776	−0.461	0.645
Cong_Imp × NFC	0.615	0.824	0.747	0.455
Model Summary (3e)	Log-Likelihood = −2103.76, AIC = 4241.51, $\text{Cond}_H = 4.7 \times 10^3$			
Cong_Imp	3.381	0.536	6.312	$< 0.001^{**}$
BRS	0.455	0.783	0.581	0.561
Cong_Imp × BRS	−0.419	0.721	−0.581	0.561
Model Summary (3f)	Log-Likelihood = −2103.81, AIC = 4241.62, $\text{Cond}_H = 4.3 \times 10^3$			

Table 7: Summary of posthoc comparison statistics for each of the interaction effects on three measures (DV: Rank Position, Recall Ability, and Information Interpretation). Inferential statistics for posthoc regression analyses (via CLMMs) and coefficient comparisons (via two-sampled Z-tests) are shown on the left and right, respectively. N represents the sample size, n_{px} denotes the number of participants included in the sample, and ∗∗ marks a significant effect.

DV	Condition	N (n_{px})	Coef. (β)	S.E.	Z	p-value	Coefficient Comparison
Rank Position	Strength (High)	634 (41)	2.305	0.405	5.698	$< 0.001^{**}$	High > Low
	Strength (Low)	345 (39)	1.196	0.408	2.931	0.003^{**}	$Z = 40.968, p < 0.001$
Rank Position	WPCS (High)	499 (21)	2.618	0.431	6.073	$< 0.001^{**}$	High > Low
	WPCS (Low)	480 (20)	1.183	0.436	2.711	0.007^{**}	$Z = 52.076, p < 0.001$
Recall Ability	CRT (High)	648 (24)	0.367	0.380	0.965	0.335	Not applicable
	CRT (Low)	331 (14)	−0.807	0.354	−2.280	0.022^{**}	
Info. Interpret.	CRT (High)	648 (27)	2.833	0.442	6.399	$< 0.001^{**}$	Low > High
	CRT (Low)	331 (14)	2.955	0.514	5.750	$< 0.001^{**}$	$Z = -4.055, p < 0.001$
Info. Interpret.	Strength (High)	634 (41)	3.894	0.408	9.533	$< 0.001^{**}$	Not Applicable
	Strength (Low)	345 (39)	0.915	0.547	1.673	0.094	
Info. Interpret.	Interest (High)	499 (21)	3.302	0.439	7.517	$< 0.001^{**}$	High > Low
	Interest (Low)	480 (20)	1.450	0.620	2.337	0.0194^{**}	$Z = 65.942, p < 0.001$
Info. Interpret.	WPCS (High)	499 (21)	3.778	0.457	8.259	$< 0.001^{**}$	High > Low
	WPCS (Low)	480 (20)	1.910	0.524	3.645	$< 0.001^{**}$	$Z = 63.858, p < 0.001$

5 Discussion

In this study, we investigated the roles of different individual and contextual factors in amplifying and moderating confirmation biases. Through three task scenarios where confirmation bias generally manifests, we exposed participants to tweet-like content items that stated a strong opinion on controversial topics. With the individual and contextual differences we collected in the study, we found that the tendency for effortful thinking, strong political beliefs, and the perceived issue strength of the tweet influence the occurrences of confirmation bias. In the remainder of this section, we discuss our findings in detail, followed by the practical and ethical implications for designing and tailoring context-aware interventions to effectively mitigate cognitive biases.

5.1 Amplifiers for Confirmation Bias

5.1.1 Content's Perceived Issue Strength. We found that the perceived tweet's issue strength interacted with the effects of ideological congruency on the information-seeking intention and information interpretation ratings. In both tasks, the effect of ideological congruency is significant when individuals interact with tweets perceived as ideologically strong, with the headline ranking task also showing that the effect of ideological congruency is stronger when the participants perceived that the tweet's issue strength was stronger. This finding implies that content perceived as a strong issue may be more likely to trigger confirmation bias. It also resonates with Zhao et al. [150], who suggested that users were more likely to share online health articles with a strong opinion stance. In addition, we found that individuals tended to recall better tweets that were perceived as ideologically stronger. This finding aligns with previous research on human memory [28, 72], showing that people tend to remember better emotionally valenced stimuli than neutral stimuli. Our results provide an empirical contribution to the confirmation bias literature as we shed light on the role of content's perceived issue strength on memory recall.

5.1.2 Individual's Political Attitudes. We found that a strong leaning towards Liberalism, reflected through high WPCS scores in our population, amplified the effects of confirmation bias on information-seeking intention and information interpretation. Notably, individuals with relatively strong liberal beliefs tended to rank higher (*i.e.*, feel inclined to read the entire tweet) headlines of tweets deemed congruent with their beliefs. They perceived it differently from those ideologically dissenting. In other words, this suggests that individuals with relatively strong political leanings may be more susceptible to using mental shortcuts and cognitive biases. Prior studies also support our findings; for example, Pennycook and Rand [103], as well as Traberg and van der Linden [139], suggested that political partisanship affects how individuals evaluate information as they perceive news with politically opposing stances or sources as less reliable. Therefore, our findings extend the prior literature: we demonstrate that individuals with strong liberal leanings are more susceptible to confirmation bias than those with neutral (or moderate conservative) beliefs.

5.1.3 Individual's Thinking Styles. Our results also highlighted that individuals with a lower tendency for effortful thinking, *i.e.*, those who scored lower on the Cognitive Reflection Test (CRT), relied significantly on ideological congruence when interpreting and recalling information. In the information interpretation scenario, we found a stronger effect size of ideological congruence when comparing individuals who scored lower on CRT and those who scored higher. Importantly, we found that the tendency for low effortful thinking determined how information is recalled: individuals with a lower effortful thinking tendency recalled better information that opposed their beliefs. This aligns with findings from Greene et al. [50], who reported that individuals with a lower effortful thinking tendency, measured similarly through CRT, formed more false memories than those with a higher tendency. However, our results contrast with Strømsø et al. [130], who found that individuals who better recalled belief-inconsistent information tended to score higher on CRT. While our work quantified the recall ability on a 4-Likert scale, they measured recall using a binary construct (*i.e.*, whether the recalled response is consistent with the original content or not). More research is needed to investigate the joint role of effortful thinking and prior beliefs in information recall.

5.1.4 Task Design and Modality. When considering all scenarios in conjunction, the factors we identified in this study amplify confirmation bias differently in each task. For example, while we found that the tweet's issue strength and the individual's effortful thinking tendency influence confirmation bias in information-seeking intention, the former did not show an interaction effect on recall ability. Similarly, the individual's topic interest only appeared as a susceptibility factor of confirmation bias in the information interpretation scenario. This finding, therefore, suggests that the *interaction context*, *i.e.*, the nature of the task, could be an influencing factor to bias susceptibility. Similarly, Vedejová and Čavojová [144] investigated confirmation bias across three scenarios (information seeking, interpretation, and recall) but did not find an effect of confirmation bias in information recall, in which the authors acknowledged that the nature of the task deployed in the study could influence how confirmation bias manifests. In the context of AI trust calibration, Ha and Kim [53] showed that different modalities of interventions (visual and textual explanation) could influence the effectiveness of confirmation bias mitigation. In psychology, Jonas et al. [65] suggested that the more natural setting of the information task leads to a stronger biased information processing, and, therefore, the choice of experimental design could influence (or confound) the occurrences of confirmation bias. In this study, however, we did not investigate the interaction effects between the contextual and individual factors on confirmation bias (*e.g.*, would effortful thinking tendency interact with the choice of task design?). Thus, we call for future research to consider multiple factors in conjunction when studying bias susceptibility.

5.2 Practical and Ethical Implications for Context-Aware Intervention Design

Synthesising these insights, our results can help inform the design of cognitive bias interventions by taking into account bias susceptibility factors, namely, the content's perceived issue strength, the user's political leaning, and thinking styles. With ideologically strong information amplifying confirmation bias, our findings suggest platforms could target content items that tend to be perceived by users as a strong stance. Consequently, platforms could soften

the issue strength of social media content using linguistic models [78, 107, 138] to detect and adjust the content's stance and sentiment towards a more nuanced perspective to help safeguard users from falling victim to their biases. It is worth noting that users may perceive the same content differently, potentially seeing it as stronger than intended by the content creator. On the other hand, intervention designers could statistically model how users perceive the issue strength of different expressions based on their rating of the content – the same measure we employed in this study (Strength) – and personalise interventions that adapt according to the content's tendency to be perceived as a firm stance. However, we acknowledge that the strength-softening measure may be viewed as a form of censorship and benevolent paternalism (*i.e.*, limiting the agency and controlling the content's stance in the *best interest* of the people). This concern is similar to the critiques about nudging as it limits users' autonomy over their decision-making [64, 132]. It may introduce novel production incentives as well as externalities [71] as content creators may divert from producing content with the potential to be demoted on the platform. On the other hand, we argue that platforms carefully consider their measures while giving users the autonomy to decide which version of the content – original or filtered – they would like to see.

Furthermore, platforms could specifically consider users' individual differences. Preventive interventions, such as psychological inoculation [84], media literacy building [52], or imposing safeguarding mechanism [32, 82], could also be targetted to specific user groups to train and fortify them against manipulation. The abovementioned bias susceptibility factors can be inferred via users' daily usage and interaction on social media platforms [100] or collected via one-off questionnaires. Nevertheless, we note that with the ability to identify individuals' bias susceptibility, platforms should leverage this data ethically not to amplify their users' cognitive biases. Importantly, by identifying biases, there is a risk of abuse to reaffirm and influence people's beliefs and decision-making. The Cambridge Analytica scandal [16, 60] is a prominent example where intimate knowledge about users' psychological traits was used to tailor targeted messaging to manipulate their opinion formation and decision-making. Therefore, the ability to determine individuals' specific tendencies and bias susceptibility should be treated with great caution. We envision that platforms could practically provide transparency of the personalised interventions through informed consent, giving users the awareness of what data are being collected, as well as an explanation of how and why a certain intervention is being tailored to them [152]. Users should always be given the autonomy to review what interventions are being applied to them, the potential impacts for them (*e.g.*, this intervention may subconsciously steer your news feed behaviour), and the ability to opt-out. We also acknowledge that data protection laws (for example, European Union's GDPR[3]) may restrict the ability to collect sensitive data which are used to inform personalised interventions.

In summary, our work paves the way towards *context-aware* interventions which adapt to the *user* and *interaction context*. The literature clearly indicates that there is no one-size-fits-all approach to mitigating cognitive biases because these cognitive tendencies manifest differently depending on the context of users, systems, and their interactions [2, 91, 110]. Cresci et al. [24] also argued that interventions could be shifted from a platform-centred approach to a personalised manner through user and context modelling. By extending the notion of context-awareness [116], we can develop computing systems that personalise not only to one's cognitive biases but, at the same time, mitigate their adverse effects using the same characteristics deduced from users' interaction data [97]. Nonetheless, it is unclear how effective personalised, context-aware interventions would be. Rieger et al. [113] investigated the effect of cognitive reflection style (collected through CRT) on the effectiveness of nudging and boosting interventions. While they did not find a significant effect, the authors argued that the effect might have been moderated by other individual and contextual factors. In line with their work, we envision that future research evaluates the effectiveness of personalised cognitive bias interventions across different populations, contexts, and task designs.

[3]https://gdpr-info.eu/issues/personal-data/

5.3 Limitations

There are several limitations to our study. First, the distribution of WPCS indicated that most of our participants reported strong liberal ideologies, while those who held conservative beliefs were small in number. This is a common phenomenon when recruiting study participants, especially from a university campus [90]. This population also tends to have higher education levels and more developed critical thinking abilities, as shown in our study participants' demographics and distribution of CRT scores. While the finding indicates that the tendency for political beliefs was an influencing factor of confirmation biases, it only gives a one-sided picture as we could only draw a comparison between strong liberals on one end and neutrals or moderate conservatives on the other end. The literature suggests that individuals with conservative beliefs could be more susceptible to misinformation and to using mental shortcuts [34, 51, 67]. Meanwhile, Ditto et al. [27] and Enders et al. [37] argued that, in the US political context, liberals are not less susceptible than conservatives. Therefore, collecting more participant samples from the conservative end would give a more complete picture of bias susceptibility, allowing a better generalisability of our findings. We also acknowledge that the number of participants in this study is limited due to its in-person setting. While our sample size (N = 42) is properly powered, we suggest future works could replicate our study through online experiments and recruit a larger, more heterogeneous set of participants.

We deployed our stimuli and tasks on Qualtrics. We acknowledge that it may not represent realistic information consumption scenarios on social media, which may consist of source cues, visual information, and social interaction with other users. Nevertheless, our main focus is on how the information is processed and how individual and contextual factors influence confirmation bias in such activities. Qualtrics, therefore, allowed us to separate confounds and closer study factors for bias susceptibility. We envision that future research further investigates bias susceptibility in higher fidelity settings, resembling real-world information consumption scenarios, while accounting for potential confounds, such as the effects of source bias and the coherence between prior beliefs and the stance of the information content [135].

We employed self-report questionnaires to gauge the tweet's perceived issue strength (Strength). This measure may be prone to subjectivity and, therefore, may not offer the best indicator for content's strong stance as applied in context-aware interventions. We recommend that future research consider using a crowd of people (that represent diverse political inclinations) to determine the issue strength of the media content.

Moreover, the recall responses and ability scores may be prone to noise. While we instructed participants to write down everything they could remember about the tweets, each participant approached this task differently. For example, some participants reported that they tried to summarise and elaborate the tweets into *pro* and *con* sides. At the same time, some listed the details of each tweet they remembered. Because the task did not restrict what the participants could write down, the richness of the recall responses depended on the participant's discretion. Some participants provided extensive recall, while some wrote only the critical aspects of each tweet. In future work, we suggest that free-recall measures could be accompanied by cued recall, *e.g.*, asking participants multiple-choice questions about the tweets or recognition tasks, where participants are asked to label items they remember.

Lastly, we used the word association test to gauge participants' ideological stances. Although it showed a strong correlation with the explicit self-assessment stance, Cong_Exp, our implicit measure may be prone to the same self-presentation [129], preference falsification issues [79], and partisan bias [13]. We suggest that future research consider measurements that better separate out these confounds, for example, the implicit association test [26, 124], to capture the nuanced strength of ideological stance and political concordances.

6 Conclusion

Cognitive biases offer useful heuristics that allow us to sift through the sheer amounts of online information quickly and effectively. At the same time, this comes at the cost of undermining the quality of our decision-making. Cognitive biases tend to be difficult to be effectively mitigated. People seem to be susceptible to acting on their biases to different degrees. In this work, we shed light on the influencing role of individual and contextual factors of cognitive biases in three scenarios: information-seeking intention, recall, and interpretation – three tasks commonly found when sifting through information online. Specifically, we investigated how these factors amplify confirmation bias – the reliance on prior beliefs – when exposed to ideologically polarised content. We found that the individual's strong political beliefs, low-effortful thinking tendency, and interest in the issue, as well as the content's perceived firm stance and the nature of the interaction with information, render users especially susceptible to confirmation bias. These insights pave the way towards more targeted safeguarding mechanisms and designing more effective, context-aware intervention systems that consider individual and contextual differences to mitigate cognitive biases, keep people safe online, and support more informed decision-making. Our findings inform measures on social media platforms to (1) reduce language that tends to be perceived as emotional or firm expressions and (2) target preventive interventions, such as safeguarding and media literacy-building mechanisms, on users with tendencies for low-effortful thinking and strong political beliefs. At the same time, designers should take these characteristics with great care and transparency, as they could open doors for paternalism and manipulation.

Acknowledgments

We thank participants in our studies and members of the HCI group at the University of Melbourne for their feedback, which helped to positively shape this paper.

References

[1] Alberto Acerbi. 2019. Cognitive attraction and online misinformation. *Palgrave Communications* 5, 1 (2019).
[2] Zhila Aghajari, Eric P. S. Baumer, and Dominic DiFranzo. 2023. Reviewing Interventions to Address Misinformation: The Need to Expand Our Vision Beyond an Individualistic Focus. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 87 (apr 2023), 34 pages. https://doi.org/10.1145/3579520
[3] Faisal Alatawi, Lu Cheng, Anique Tahir, Mansooreh Karami, Bohan Jiang, Tyler Black, and Huan Liu. 2021. A Survey on Echo Chambers on Social Media: Description, Detection and Mitigation. *arXiv preprint arXiv:2112.05084* (2021). https://arxiv.org/abs/2112.05084
[4] Jennifer Allen, Cameron Martel, and David G Rand. 2022. Birds of a Feather Don't Fact-Check Each Other: Partisanship and the Evaluation of News in Twitter's Birdwatch Crowdsourced Fact-Checking Program. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 245, 19 pages. https://doi.org/10.1145/3491102.3502040
[5] Leif Azzopardi. 2021. Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) *(CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 27–37. https://doi.org/10.1145/3406522.3446023
[6] William Badke. 2018. Fake news, confirmation bias, the search for truth, and the theology student. *Theological Librarianship* 11, 2 (2018), 4–7. https://doi.org/10.31046/tl.v11i2.519
[7] Ricardo Baeza-Yates. 2018. Bias on the Web. *Commun. ACM* 61, 6 (May 2018), 54–61. https://doi.org/10.1145/3209581
[8] Bence Bago, David G Rand, and Gordon Pennycook. 2020. Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of experimental psychology: general* 149, 8 (2020), 1608.
[9] Alexander Benlian. 2015. Web Personalization Cues and Their Differential Effects on User Assessments of Website Value. *Journal of Management Information Systems* 32, 1 (2015), 225–260. https://doi.org/10.1080/07421222.2015.1029394
[10] H. Berghel. 2018. Malice Domestic: The Cambridge Analytica Dystopia. *Computer* 51, 05 (may 2018), 84–89. https://doi.org/10.1109/MC.2018.2381135
[11] Nattapat Boonprakong, Xiuge Chen, Catherine Davey, Benjamin Tag, and Tilman Dingler. 2023. Bias-Aware Systems: Exploring Indicators for the Occurrences of Cognitive Biases When Facing Different Opinions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 27, 19 pages. https://doi.org/10.1145/3544548.3580917
[12] Nattapat Boonprakong, Benjamin Tag, and Tilman Dingler. 2023. Designing Technologies to Support Critical Thinking in an Age of Misinformation. *IEEE Pervasive Computing* (2023), 1–10. https://doi.org/10.1109/MPRV.2023.3275514
[13] John G. Bullock and Gabriel Lenz. 2019. Partisan Bias in Surveys. *Annual Review of Political Science* 22, Volume 22, 2019 (2019), 325–342. https://doi.org/10.1146/annurev-polisci-051117-050904
[14] Pavlo Burda, Luca Allodi, and Nicola Zannone. 2024. Cognition in Social Engineering Empirical Research: A Systematic Literature Review. *ACM Trans. Comput.-Hum. Interact.* 31, 2 (2024). https://doi.org/10.1145/3635149
[15] John T Cacioppo and Richard E Petty. 1982. The Need for Cognition. *Journal of personality and social psychology* 42, 1 (1982), 116.
[16] Carole Cadwalladr. 2017. The great British Brexit robbery: how our democracy was hijacked. *The Guardian* 7 (2017).
[17] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300733
[18] Sijing Chen, Lu Xiao, and Akit Kumar. 2023. Spread of misinformation on social media: What contributes to it and how to combat it. *Computers in Human Behavior* 141 (2023), 107643. https://doi.org/10.1016/j.chb.2022.107643

[19] Rune Haubo B Christensen. 2015. A Tutorial on fitting Cumulative Link Models with the ordinal Package. *Retrieved from www.cran.r-project.org/package=ordinal* (2015).

[20] Russ Clay, Jessica M. Barber, and Natalie J. Shook. 2013. Techniques for Measuring Selective Exposure: A Critical Review. *Communication Methods and Measures* 7, 3-4 (2013), 147–171. https://doi.org/10.1080/19312458.2013.813925 arXiv:https://doi.org/10.1080/19312458.2013.813925

[21] Katherine Clayton, Jase Davis, Kristen Hinckley, and Yusaku Horiuchi. 2019. Partisan motivated reasoning and misinformation in the media: Is news from ideologically uncongenial sources more suspicious? *Japanese Journal of Political Science* 20, 3 (2019), 129–142. https://doi.org/10.1017/S1468109919000082

[22] Clifford C. Clogg, Eva Petkova, and Adamantios Haritou. 1995. Statistical Methods for Comparing Regression Coefficients Between Models. *Amer. J. Sociology* 100, 5 (1995), 1261–1293. https://doi.org/10.1086/230638 arXiv:https://doi.org/10.1086/230638

[23] Sonia Correa. 2003. Abortion is a global political issue. In *A DAWN Supplement for the World Social Forum, Porto Alegre, 23–28 January 2003.*

[24] Stefano Cresci, Amaury Trujillo, and Tiziano Fagni. 2022. Personalized Interventions for Online Moderation. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media* (Barcelona, Spain) *(HT '22)*. Association for Computing Machinery, New York, NY, USA, 248–251. https://doi.org/10.1145/3511095.3536369

[25] Kristen D. Deppe, Frank J. Gonzalez, Jayme L. Neiman, Carly Jacobs, Jackson Pahlke, Kevin B. Smith, and John R. Hibbing. 2015. Reflective liberals and intuitive conservatives: A look at the Cognitive Reflection Test and ideology. *Judgment and Decision Making* 10, 4 (2015), 314–331. https://doi.org/10.1017/S1930297500005131

[26] Tilman Dingler, Benjamin Tag, David A. Eccles, Niels van Berkel, and Vassilis Kostakos. 2022. Method for Appropriating the Brief Implicit Association Test to Elicit Biases in Users. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 243, 16 pages. https://doi.org/10.1145/3491102.3517570

[27] Peter H. Ditto, Brittany S. Liu, Cory J. Clark, Sean P. Wojcik, Eric E. Chen, Rebecca H. Grady, Jared B. Celniker, and Joanne F. Zinger. 2019. At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives. *Perspectives on Psychological Science* 14, 2 (2019), 273–291. https://doi.org/10.1177/1745691617746796 arXiv:https://doi.org/10.1177/1745691617746796 PMID: 29851554.

[28] Florin Dolcos, Kevin S LaBar, and Roberto Cabeza. 2006. The memory enhancing effect of emotion: Functional neuroimaging evidence. (2006).

[29] 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* (Virtual Event, Canada) *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 295–305. https://doi.org/10.1145/3404835.3462851

[30] James N. Druckman. 2012. The Politics of Motivation. *Critical Review: A Journal of Politics and Society* 24, 2 (2012), 199–216. https://doi.org/10.1080/08913811.2012.711022

[31] AH Eagly, S Chen, S Chaiken, and K Shaw-Barnes. 1999. The impact of attitudes on memory: An affair to remember. *PSYCHOLOGICAL BULLETIN* 125, 1 (JAN 1999), 64–89. https://doi.org/10.1037/0033-2909.125.1.64

[32] David A Eccles, Sherah Kurnia, Tilman Dingler, and Nicholas Geard. 2021. Three Preventative Interventions to Address the Fake News Phenomenon on Social Media. In *ACIS 2021 Proceedings*, Vol. 51. https://aisel.aisnet.org/acis2021/51

[33] Ullrich KH Ecker, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K Fazio, Nadia Brashier, Panayiota Kendeou, Emily K Vraga, and Michelle A Amazeen. 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology* 1, 1 (2022), 13–29.

[34] Ullrich K. H. Ecker and Li Chang Ang. 2019. Political Attitudes and the Processing of Misinformation Corrections. *Political Psychology* 40, 2 (2019), 241–260. https://doi.org/10.1111/pops.12494

[35] Ullrich K. H. Ecker, Stephan Lewandowsky, Olivia Fenton, and Kelsey Martin. 2014. Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition* 42, 2 (Feb. 2014), 292–304. https://doi.org/10.3758/s13421-013-0358-x

[36] Scott Eidelman, Christian S. Crandall, Jeffrey A. Goodman, and John C. Blanchar. 2012. Low-Effort Thought Promotes Political Conservatism. *Personality and Social Psychology Bulletin* 38, 6 (2012), 808–820. https://doi.org/10.1177/0146167212439213 arXiv:https://doi.org/10.1177/0146167212439213 PMID: 22427384.

[37] Adam Enders, Christina Farhart, Joanne Miller, Joseph Uscinski, Kyle Saunders, and Hugo Drochon. 2023. Are Republicans and Conservatives More Likely to Believe Conspiracy Theories? *Political Behavior* 45, 4 (Dec. 2023), 2001–2024. https://doi.org/10.1007/s11109-022-09812-3

[38] Jonathan St BT Evans. 1989. *Bias in human reasoning: Causes and consequences.* Lawrence Erlbaum Associates, Inc.

[39] Jonathan St. B. T. Evans. 2008. Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology* 59, 1 (2008), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629 arXiv:https://doi.org/10.1146/annurev.psych.59.103006.093629 PMID: 18154502.

[40] Jonathan St. B. T. Evans, Simon J. Handley, Helen Neilens, and David Over. 2010. The influence of cognitive ability and instructional set on causal conditional inference. *Quarterly Journal of Experimental Psychology* 63, 5 (2010), 892–909. https://doi.org/10.1080/17470210903111821 arXiv:https://doi.org/10.1080/17470210903111821 PMID: 19728225.

[41] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41, 4 (2009), 1149–1160.

[42] Leon Festinger. 1962. Cognitive Dissonance. *Scientific American* 207, 4 (1962), 93–106. http://www.jstor.org/stable/24936719

[43] Peter Fischer, Eva Jonas, Dieter Frey, and Stefan Schulz-Hardt. 2005. Selective exposure to information: The impact of information limits. *European Journal of social psychology* 35, 4 (2005), 469–492.

[44] Shane Frederick. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19, 4 (December 2005), 25–42. https://doi.org/10.1257/089533005775196732

[45] Peter Frost, Bridgette Casey, Kaydee Griffin, Luis Raymundo, Christopher Farrell, and Ryan Carrigan. 2015. The Influence of Confirmation Bias on Memory and Source Monitoring. *The Journal of General Psychology* 142, 4 (2015), 238–252. https://doi.org/10.1080/00221309.2015.1084987 arXiv:https://doi.org/10.1080/00221309.2015.1084987 PMID: 26649923.

[46] R. Kelly Garrett. 2009. Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate. *Journal of Communication* 59, 4 (12 2009), 676–699. https://doi.org/10.1111/j.1460-2466.2009.01452.x arXiv:https://academic.oup.com/joc/article-pdf/59/4/676/22323928/jjnlcom0676.pdf

[47] Christine Geeng, Savanna Yee, and Franziska Roesner. 2020. Fake News on Facebook and Twitter: Investigating How People (Don't) Investigate. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376784

[48] Gerd Gigerenzer. 2004. *Fast and Frugal Heuristics: The Tools of Bounded Rationality.* John Wiley & Sons, Ltd, Chapter 4, 62–88. https://doi.org/10.1002/9780470752937.ch4 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470752937.ch4

[49] Eduardo Graells-Garrido, Mounia Lalmas, and Ricardo Baeza-Yates. 2016. Data Portraits and Intermediary Topics: Encouraging Exploration of Politically Diverse Profiles. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 228–240. https://doi.org/10.1145/2856767.2856776

[50] Ciara M. Greene, Robert A. Nash, and Gillian Murphy. 2021. Misremembering Brexit: partisan bias and individual predictors of false memories for fake news stories among Brexit voters. *Memory* 29, 5 (2021), 587–604. https://doi.org/10.1080/09658211.2021.1923754 arXiv:https://doi.org/10.1080/09658211.2021.1923754 PMID: 33971789.

[51] Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances* 5, 1 (2019), eaau4586. https://doi.org/10.1126/sciadv.aau4586 arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.aau4586

[52] Andrew M. Guess, Michael Lerner, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. 2020. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences* 117, 27 (2020), 15536–15545. https://doi.org/10.1073/pnas.1920498117 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1920498117

[53] Taehyun Ha and Sangyeon Kim. 2023. Improving Trust in AI with Mitigating Confirmation Bias: Effects of Explanation Type and Debiasing Strategy for Decision-Making with Explainable AI. *International Journal of Human–Computer Interaction* 0, 0 (2023), 1–12. https://doi.org/10.1080/10447318.2023.2285640 arXiv:https://doi.org/10.1080/10447318.2023.2285640

[54] Ulrike Hahn and Adam J.L. Harris. 2014. Chapter Two - What Does It Mean to be Biased: Motivated Reasoning and Rationality. Psychology of Learning and Motivation, Vol. 61. Academic Press, 41–102. https://doi.org/10.1016/B978-0-12-800283-4.00002-2

[55] Martie G. Haselton, Gregory A. Bryant, Andreas Wilke, David A. Frederick, Andrew Galperin, Willem E. Frankenhuis, and Tyler Moore. 2009. Adaptive Rationality: An Evolutionary Perspective on Cognitive Bias. *Social Cognition* 27, 5 (2009), 733–763. https://doi.org/10.1521/soco.2009.27.5.733 arXiv:https://doi.org/10.1521/soco.2009.27.5.733

[56] Martie G Haselton, Daniel Nettle, and Damian R Murray. 2015. The evolution of cognitive bias. *The handbook of evolutionary psychology* (2015), 1–20.

[57] Dallas Havens. 2022. *A Quick Look: The Debate Surrounding Ethical Vegetarianism.* https://dallashavens.wordpress.com/2022/07/08/a-quick-look-the-debate-

surrounding-ethical-vegetarianism/
[58] J.P. Henningham. 1996. A 12-item scale of social conservatism. *Personality and Individual Differences* 20, 4 (1996), 517–519. https://doi.org/10.1016/0191-8869(95)00192-1
[59] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 48 (may 2020), 27 pages. https://doi.org/10.1145/3392854
[60] Jim Isaak and Mina J. Hanna. 2018. User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer* 51, 8 (2018), 56–59. https://doi.org/10.1109/MC.2018.3191268
[61] Kaixin Ji. 2023. Quantifying and Measuring Confirmation Bias in Information Retrieval Using Sensors. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* (Cancun, Quintana Roo, Mexico) *(UbiComp/ISWC '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 236–240. https://doi.org/10.1145/3594739.3610765
[62] Ángel V Jiménez, Alex Mesoudi, and Jamshid J Tehrani. 2020. No evidence that omission and confirmation biases affect the perception and recall of vaccine-related information. *PloS one* 15, 3 (2020), e0228898. https://doi.org/10.1371/journal.pone.0228898
[63] Blair T Johnson and Alice H Eagly. 1989. Effects of involvement on persuasion: A meta-analysis. *Psychological bulletin* 106, 2 (1989), 290.
[64] Christine Jolls and Cass R. Sunstein. 2006. Debiasing through Law. *The Journal of Legal Studies* 35, 1 (2006), 199–242. https://doi.org/10.1086/500096 arXiv:https://doi.org/10.1086/500096
[65] Eva Jonas, Stefan Schulz-Hardt, Dieter Frey, and Norman Thelen. 2001. Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. *Journal of personality and social psychology* 80, 4 (2001), 557.
[66] Kristyn A. Jones, William E. Crozier, and Deryn Strange. 2017. Believing is Seeing: Biased Viewing of Body-Worn Camera Footage. *Journal of Applied Research in Memory and Cognition* 6, 4 (2017), 460–474. https://doi.org/10.1016/j.jarmac.2017.07.007
[67] John T Jost, Sander van der Linden, Costas Panagopoulos, and Curtis D Hardin. 2018. Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Current Opinion in Psychology* 23 (2018), 77–83. https://doi.org/10.1016/j.copsyc.2018.01.003 Shared Reality.
[68] Dan M Kahan. 2015. The politically motivated reasoning paradigm. *Emerging Trends in Social & Behavioral Sciences, Forthcoming* (2015).
[69] Daniel Kahneman. 2011. *Thinking, Fast and Slow.* Macmillan.
[70] Robert A. Kaufman, Michael Robert Haupt, and Steven P. Dow. 2022. Who's in the Crowd Matters: Cognitive Factors and Beliefs Predict Misinformation Assessment Accuracy. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 553 (nov 2022), 18 pages. https://doi.org/10.1145/3555611
[71] Devansh Kaushik. 2024. Policy Responses To Fake News On Social Media Platforms: A Law And Economics Analysis. *Statute Law Review* 45, 1 (02 2024), hmae013. https://doi.org/10.1093/slr/hmae013 arXiv:https://academic.oup.com/slr/article-pdf/45/1/hmae013/56770687/hmae013.pdf
[72] Elizabeth A Kensinger and Suzanne Corkin. 2003. Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & cognition* 31, 8 (2003), 1169–1180.
[73] Paul A. Klaczynski. 2000. Motivated Scientific Reasoning Biases, Epistemological Beliefs, and Theory Polarization: A Two-Process Approach to Adolescent Cognition. *Child Development* 71, 5 (2000), 1347–1366. https://doi.org/10.1111/1467-8624.00232
[74] Joseph T Klapper. 1960. The effects of mass communication. (1960).
[75] Jan Kleinnijenhuis. 2008. *Negativity.* John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405186407.wbiecn005
[76] Keiichi Kobayashi. 2010. Strategic Use of Multiple Texts for the Evaluation of Arguments. *Reading Psychology* 31, 2 (2010), 121–149. https://doi.org/10.1080/02702710902754192 arXiv:https://doi.org/10.1080/02702710902754192
[77] Jon A Krosnick, Charles M Judd, and Bernd Wittenbrink. 2005. The measurement of attitudes. *The handbook of attitudes* 21 (2005), 76.
[78] Dilek Küçük and Fazli Can. 2020. Stance Detection: A Survey. *ACM Comput. Surv.* 53, 1, Article 12 (feb 2020), 37 pages. https://doi.org/10.1145/3369026
[79] Timur Kuran. 1997. *Private truths, public lies: The social consequences of preference falsification.* Harvard University Press.
[80] David Lazer, Matthew Baum, Nir Grinberg, Lisa Friedland, Kenneth Joseph, Will Hobbs, and Carolina Mattsson. 2017. Combating fake news: An agenda for research and action. (2017).
[81] Colin Lescarret, Valérie Le Floch, Jean-Christophe Sakdavong, Jean-Michel Boucheix, André Tricot, and Franck Amadieu. 2023. The impact of students' prior attitude on the processing of conflicting videos: a comparison between middle-school and undergraduate students. *European Journal of Psychology of Education* 38, 2 (June 2023), 519–544. https://doi.org/10.1007/s10212-022-00634-9
[82] Stephan Lewandowsky, Ullrich K.H. Ecker, and John Cook. 2017. Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition* 6, 4 (2017), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008
[83] Stephan Lewandowsky, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. 2012. Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest* 13, 3 (2012), 106–131. https://doi.org/10.1177/1529100612451018 arXiv:https://doi.org/10.1177/1529100612451018 PMID: 26173286.
[84] Stephan Lewandowsky and Sander van der Linden. 2021. Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology* 32, 2 (2021), 348–384. https://doi.org/10.1080/10463283.2021.1876983 arXiv:https://doi.org/10.1080/10463283.2021.1876983
[85] Ming-Hui Li, Zhiqin Chen, and Li-Lin Rao. 2022. Emotion, analytic thinking and susceptibility to misinformation during the COVID-19 outbreak. *Computers in Human Behavior* 133 (2022), 107295. https://doi.org/10.1016/j.chb.2022.107295
[86] Q. Vera Liao and Wai-Tat Fu. 2013. Beyond the Filter Bubble: Interactive Effects of Perceived Threat and Topic Involvement on Selective Exposure to Information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2359–2368. https://doi.org/10.1145/2470654.2481326
[87] Q. Vera Liao, Wai-Tat Fu, and Sri Shilpa Mamidi. 2015. It Is All About Perspective: An Exploration of Mitigating Selective Exposure with Aspect Indicators. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1439–1448. https://doi.org/10.1145/2702123.2702570
[88] Falk Lieder, Thomas L Griffiths, and Ming Hsu. 2018. Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological review* 125, 1 (2018), 1.
[89] Scott O. Lilienfeld, Rachel Ammirati, and Kristin Landfield. 2009. Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare? *Perspectives on Psychological Science* 4, 4 (2009), 390–398. https://doi.org/10.1111/j.1745-6924.2009.01144.x arXiv:https://doi.org/10.1111/j.1745-6924.2009.01144.x PMID: 26158987.
[90] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 143, 14 pages. https://doi.org/10.1145/3411764.3445488
[91] Jiqun Liu. 2023. Toward A Two-Sided Fairness Framework in Search and Recommendation. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 236–246. https://doi.org/10.1145/3576840.3578332
[92] Elizabeth F. Loftus and John C. Palmer. 1974. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior* 13, 5 (1974), 585–589. https://doi.org/10.1016/S0022-5371(74)80011-3
[93] 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, 11 (2020), 1102–1109. https://doi.org/10.1038/s41562-020-0889-7
[94] Cameron Martel, Gordon Pennycook, and David G. Rand. 2020. Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications* 5, 1 (Oct. 2020), 47. https://doi.org/10.1186/s41235-020-00252-3
[95] Gillian Murphy, Emma Murray, and Doireann Gough. 2021. Attitudes towards feminism predict susceptibility to feminism-related fake news. *Applied Cognitive Psychology* 35, 5 (2021), 1182–1192. https://doi.org/10.1002/acp.3851 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.3851
[96] Raymond S Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology* 2, 2 (1998), 175–220.
[97] Alexander Nussbaumer, Katrien Verbert, Eva-Catherine Hillemann, Michael A Bedek, and Dietrich Albert. 2016. A framework for cognitive bias detection and feedback in a visual analytics environment. In *2016 European Intelligence and Security Informatics Conference (EISIC)*. IEEE, 148–151.
[98] Aileen Oeberst and Roland Imhoff. 2023. Toward Parsimony in Bias Research: A Proposed Common Framework of Belief-Consistent Information Processing for a Set of Biases. *Perspectives on Psychological Science* 18, 6 (2023), 1464–1487. https://doi.org/10.1177/17456916221148147 arXiv:https://doi.org/10.1177/17456916221148147 PMID: 36930530.
[99] Margit E Oswald and Stefan Grosjean. 2004. Confirmation bias. *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* 79 (2004), 83.
[100] Irene V Pasquetto, Briony Swire-Thompson, Michelle A Amazeen, Fabrício Benevenuto, Nadia M Brashier, Robert M Bond, Lia C Bozarth, Ceren Budak, Ullrich KH Ecker, Lisa K Fazio, et al. 2020. Tackling misinformation: What researchers could do with social media data. *The Harvard Kennedy School Misinformation Review* (2020).

[101] David Paternotte. 2015. Global times, global debates? Same-sex marriage worldwide. *Social Politics: International Studies in Gender, State & Society* 22, 4 (2015), 653–674.

[102] Gordon Pennycook, James Allan Cheyne, Nathaniel Barr, Derek J. Koehler, and Jonathan A. Fugelsang. 2015. On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making* 10, 6 (2015), 549–563. https://doi.org/10.1017/S1930297500006999

[103] Gordon Pennycook and David G. Rand. 2019. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188 (2019), 39–50. https://doi.org/10.1016/j.cognition.2018.06.011 The Cognitive Science of Political Thought.

[104] Gordon Pennycook and David G. Rand. 2020. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality* 88, 2 (2020), 185–200. https://doi.org/10.1111/jopy.12476

[105] Gordon Pennycook and David G. Rand. 2021. The Psychology of Fake News. *Trends in Cognitive Sciences* 25, 5 (May 2021), 388–402. https://doi.org/10.1016/j.tics.2021.02.007 Publisher: Elsevier.

[106] Suppanut Pothirattanachaikul, Takehiro Yamamoto, Yusuke Yamamoto, and Masatoshi Yoshikawa. 2019. Analyzing the Effects of Document's Opinion and Credibility on Search Behaviors and Belief Dynamics. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (Beijing, China) *(CIKM '19)*. Association for Computing Machinery, New York, NY, USA, 1653–1662. https://doi.org/10.1145/3357384.3357886

[107] Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1650–1659.

[108] Tobias Richter. 2015. Validation and Comprehension of Text Information: Two Sides of the Same Coin. *Discourse Processes* 52, 5-6 (2015), 337–355. https://doi.org/10.1080/0163853X.2015.1025665

[109] Tobias Richter and Johanna Maier. 2017. Comprehension of Multiple Documents With Conflicting Information: A Two-Step Model of Validation. *Educational Psychologist* 52, 3 (2017), 148–166. https://doi.org/10.1080/00461520.2017.1322968

[110] Alisa Rieger. 2022. Interactive Interventions to Mitigate Cognitive Bias. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (Barcelona, Spain) *(UMAP '22)*. Association for Computing Machinery, New York, NY, USA, 316–320. https://doi.org/10.1145/3503252.3534362

[111] 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* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 248, 8 pages. https://doi.org/10.1145/3544549.3585693

[112] Alisa Rieger, Tim Draws, Mariët Theune, and Nava Tintarev. 2021. This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media* (Virtual Event, USA) *(HT '21)*. Association for Computing Machinery, New York, NY, USA, 189–199. https://doi.org/10.1145/3465336.3475101

[113] 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 Trans. Web* (nov 2023). https://doi.org/10.1145/3635034 Just Accepted.

[114] Alisa Rieger, Mariët Theune, and Nava Tintarev. 2020. Toward Natural Language Mitigation Strategies for Cognitive Biases in Recommender Systems. In *2nd Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence*. Association for Computational Linguistics, Dublin, Ireland, 50–54. https://aclanthology.org/2020.nl4xai-1.11

[115] Emily Saltz, Claire R Leibowicz, and Claire Wardle. 2021. Encounters with Visual Misinformation and Labels Across Platforms: An Interview and Diary Study to Inform Ecosystem Approaches to Misinformation Interventions. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 340, 6 pages. https://doi.org/10.1145/3411763.3451807

[116] B. Schilit, N. Adams, and R. Want. 1994. Context-Aware Computing Applications. In *1994 First Workshop on Mobile Computing Systems and Applications*. 85–90. https://doi.org/10.1109/WMCSA.1994.16

[117] Norbert Schwarz, Eryn Newman, and William Leach. 2016. Making the Truth Stick & the Myths Fade: Lessons from Cognitive Psychology. *Behavioral Science & Policy* 2, 1 (2016), 85–95. https://doi.org/10.1177/237946151600200110

[118] Christina Schwind and Jürgen Buder. 2012. Reducing confirmation bias and evaluation bias: When are preference-inconsistent recommendations effective – and when not? *Computers in Human Behavior* 28, 6 (2012), 2280–2290. https://doi.org/10.1016/j.chb.2012.06.035

[119] Li Shi, Nilavra Bhattacharya, Anubrata Das, and Jacek Gwizdka. 2023. True or False? Cognitive Load When Reading COVID-19 News Headlines: An Eye-Tracking Study. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 107–116. https://doi.org/10.1145/3576840.3578290

[120] Natalie J. Shook and Russell H. Fazio. 2009. Political ideology, exploration of novel stimuli, and attitude formation. *Journal of Experimental Social Psychology* 45, 4 (2009), 995–998. https://doi.org/10.1016/j.jesp.2009.04.003

[121] Herbert A Simon. 1957. A behavioral model of rational choice. *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting* (1957), 241–260.

[122] Miroslav Sirota and Marie Juanchich. 2018. Effect of response format on cognitive reflection: Validating a two-and four-option multiple choice question version of the Cognitive Reflection Test. *Behavior research methods* 50 (2018), 2511–2522. https://doi.org/10.3758/s13428-018-1029-4

[123] Brendan Spillane, Séamus Lawless, and Vincent Wade. 2017. Perception of Bias: The Impact of User Characteristics, Website Design and Technical Features. In *Proceedings of the International Conference on Web Intelligence* (Leipzig, Germany) *(WI '17)*. Association for Computing Machinery, New York, NY, USA, 227–236. https://doi.org/10.1145/3106426.3106474

[124] N. Sriram and Anthony G. Greenwald. 2009. The Brief Implicit Association Test. *Experimental Psychology* 56 (2009), 283–294. Issue 4. https://doi.org/10.1027/1618-3169.56.4.283

[125] Namrata Srivastava, Rajiv Jain, Jennifer Healey, Zoya Bylinskii, and Tilman Dingler. 2021. Mitigating the Effects of Reading Interruptions by Providing Reviews and Previews. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 229, 6 pages. https://doi.org/10.1145/3411763.3451610

[126] C STANGOR and D MCMILLAN. 1992. MEMORY FOR EXPECTANCY-CONGRUENT AND EXPECTANCY-INCONGRUENT INFORMATION - A REVIEW OF THE SOCIAL AND SOCIAL DEVELOPMENTAL LITERATURES. *PSYCHOLOGICAL BULLETIN* 111, 1 (JAN 1992), 42–61. https://doi.org/10.1037/0033-2909.111.1.42

[127] Keith E Stanovich. 1999. *Who is Rational?: Studies of Individual Differences in Reasoning*. Psychology Press.

[128] K. E. Stanovich, R. F. West, and R. Hertwig. 2000. Individual Differences in Reasoning: Implications for the Rationality Debate? *Behavioral and Brain Sciences* 23, 5 (2000), 678–678.

[129] Natalie Jomini Stroud. 2008. Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure. *Political Behavior* 30, 3 (2008), 341–366. http://www.jstor.org/stable/40213321

[130] Helge I. Strømsø, Ivar Bråten, and Tonje Stenseth. 2017. The role of students' prior topic beliefs in recall and evaluation of information from texts on socio-scientific issues. *Nordic Psychology* 69, 3 (2017), 127–142. https://doi.org/10.1080/19012276.2016.1198270 arXiv:https://doi.org/10.1080/19012276.2016.1198270

[131] Michael Sülflow, Svenja Schäfer, and Stephan Winter. 2019. Selective attention in the news feed: An eye-tracking study on the perception and selection of political news posts on Facebook. *new media & society* 21, 1 (2019), 168–190.

[132] Cass R. Sunstein and Richard H. Thaler. 2003. Libertarian Paternalism Is Not an Oxymoron. *The University of Chicago Law Review* 70, 4 (2003), 1159–1202. http://www.jstor.org/stable/1600573

[133] Charles S. Taber and Milton Lodge. 2006. Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science* 50, 3 (2006), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5907.2006.00214.x

[134] Yuko Tanaka, Miwa Inuzuka, Hiromi Arai, Yoichi Takahashi, Minao Kukita, and Kentaro Inui. 2023. Who Does Not Benefit from Fact-Checking Websites? A Psychological Characteristic Predicts the Selective Avoidance of Clicking Uncongenial Facts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 664, 17 pages. https://doi.org/10.1145/3544548.3580826

[135] Ben M Tappin, Gordon Pennycook, and David G Rand. 2020. Thinking clearly about causal inferences of politically motivated reasoning: why paradigmatic study designs often undermine causal inference. *Current Opinion in Behavioral Sciences* 34 (2020), 81–87. https://doi.org/10.1016/j.cobeha.2020.01.003 Political Ideologies.

[136] Predrag Teovanović, Goran Knežević, and Lazar Stankov. 2015. Individual differences in cognitive biases: Evidence against one-factor theory of rationality. *Intelligence* 50 (2015), 75–86. https://doi.org/10.1016/j.intell.2015.02.008

[137] R H Thaler and C R Sunstein. 2009. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Penguin Publishing Group.

[138] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2012. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology* 63, 1 (2012), 163–173. https://doi.org/10.1002/asi.21662 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.21662

[139] Cecilie Steenbuch Traberg and Sander van der Linden. 2022. Birds of a feather are persuaded together: Perceived source credibility mediates the effect of political

bias on misinformation susceptibility. *Personality and Individual Differences* 185 (2022), 111269. https://doi.org/10.1016/j.paid.2021.111269

[140] Amos Tversky and Daniel Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5, 2 (1973), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

[141] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (1974), 1124–1131.

[142] Jan-Willem van Prooijen, André P. M. Krouwel, and Thomas V. Pollet. 2015. Political Extremism Predicts Belief in Conspiracy Theories. *Social Psychological and Personality Science* 6, 5 (2015), 570–578. https://doi.org/10.1177/1948550614567356 arXiv:https://doi.org/10.1177/1948550614567356

[143] 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

[144] Dáša Vedejová and Vladimíra Čavojová. 2022. Confirmation bias in information search, interpretation, and memory recall: evidence from reasoning about four controversial topics. *Thinking & Reasoning* 28, 1 (2022), 1–28. https://doi.org/10.1080/13546783.2021.1891967 arXiv:https://doi.org/10.1080/13546783.2021.1891967

[145] Joseph A. Vitriol, Joseph Sandor, Robert Vidigal, and Christina Farhart. 2023. On the Independent Roles of Cognitive & Political Sophistication: Variation Across Attitudinal Objects. *Applied Cognitive Psychology* 37, 2 (2023), 319–331. https://doi.org/10.1002/acp.4022 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.4022

[146] Ben Wang and Jiqun Liu. 2024. Cognitively Biased Users Interacting with Algorithmically Biased Results in Whole-Session Search on Debated Topics. In *Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval* (Washington DC, USA) *(ICTIR '24)*. Association for Computing Machinery, New York, NY, USA, 227–237. https://doi.org/10.1145/3664190.3672520

[147] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300831

[148] P. C. Wason. 1960. On the Failure to Eliminate Hypotheses in a Conceptual Task. *Quarterly Journal of Experimental Psychology* 12, 3 (1960), 129–140. https://doi.org/10.1080/17470216008416717 arXiv:https://doi.org/10.1080/17470216008416717

[149] Magdalena Wischnewski, Rebecca Bernemann, Thao Ngo, and Nicole Krämer. 2021. Disagree? You Must Be a Bot! How Beliefs Shape Twitter Profile Perceptions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 160, 11 pages. https://doi.org/10.1145/3411764.3445109

[150] Haiping Zhao, Shaoxiong Fu, and Xiaoyu Chen. 2020. Promoting users' intention to share online health articles on social media: The role of confirmation bias. *Information Processing & Management* 57, 6 (2020), 102354. https://doi.org/10.1016/j.ipm.2020.102354

[151] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 436, 20 pages. https://doi.org/10.1145/3544548.3581318

[152] Qian Zhu, Leo Yu-Ho Lo, Meng Xia, Zixin Chen, and Xiaojuan Ma. 2022. Bias-Aware Design for Informed Decisions: Raising Awareness of Self-Selection Bias in User Ratings and Reviews. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 496 (nov 2022), 31 pages. https://doi.org/10.1145/3555597