How Do HCI Researchers Study Cognitive Biases? A Scoping Review

Boonprakong, Nattapat; Tag, Benjamin; Goncalves, Jorge; Dingler, Tilman

DOI
10.1145/3706598.3713450

Publication date
2025

Document Version
Final published version

Published in
CHI 2025 - Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems

How Do HCI Researchers Study Cognitive Biases? A Scoping Review

Nattapat Boonprakong
School of Computing and Information Systems
University of Melbourne
Parkville, Victoria, Australia
nboonprakong@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

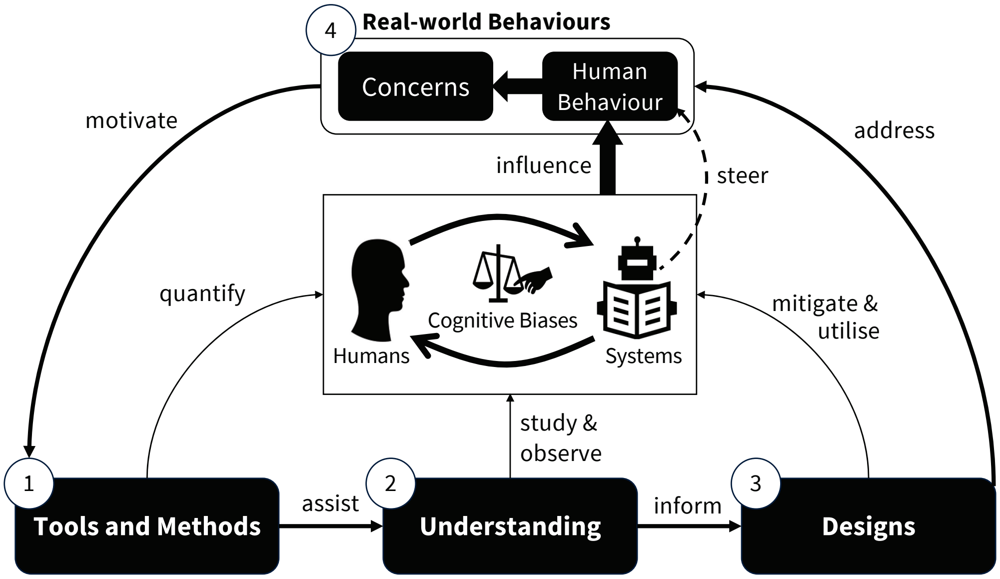

Figure 1: A summary of how HCI researchers study cognitive biases. Computing systems can trigger cognitive biases in humans and influence (or steer) their behaviours and decision-making. Cognitive biases affect the real-world behaviours of humans, which motivates HCI researchers to develop (1) tools and methods measuring the occurrences of cognitive biases to study their effects on the interaction between humans and computers. Consequently, (2) the understanding of cognitive biases informs (3) the design of computing systems, which mitigates or utilises cognitive biases and helps address (4) the real-world behaviour of humans.

Abstract

Computing systems are increasingly designed to adapt to users' cognitive states and mental models. Yet, cognitive biases affect how humans form such models and, therefore, they can impact their interactions with computers. To better understand this interplay, we conducted a scoping review to chart how Human-Computer Interaction (HCI) researchers study cognitive biases. Our findings show that computing systems not only have the potential to induce and amplify cognitive biases but also can be designed to steer users' behaviour and decision-making by capitalising on biases. We describe how HCI researchers develop algorithms and sensing methods to detect and quantify the effects of cognitive biases and discuss how we can use their understanding to inform system design. In this paper, we outline a research agenda for more theory-grounded research and highlight ethical issues when researching and designing computing systems with cognitive biases in mind as they affect real-world behaviour.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713450

CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**.

Keywords

cognitive bias; decision-making; bias-aware systems

ACM Reference Format:
Nattapat Boonprakong, Benjamin Tag, Jorge Goncalves, and Tilman Dingler. 2025. How Do HCI Researchers Study Cognitive Biases? A Scoping Review. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3706598.3713450

1 Introduction

When interacting with computers, humans form mental models – internal representations of the external reality – based on what they believe, prefer, and are familiar with [30]. The design of everyday user interfaces, such as desktops, digital games, or online websites, is predominantly based on the mental representation of humans from real-world physical objects. However, human mental models are subject to bounded rationality [166]. Humans use simple rules of thumb, developed through their beliefs and experience of the world, to sift through the complexity of everyday decision-making. The pioneers of behavioural economics, Tversky and Kahneman [97, 187, 188] coined such phenomena as **cognitive bias** and documented different ways such biases systematically skew human behaviours and judgements. For example, the anchoring bias makes us tend to rather stick with the first piece of information we encounter [188], or the framing effect influences people to make decisions differently based on how the choices are presented [189]. While computing systems are built to adapt to people's cognitive states and mental models [22, 30], cognitive biases affect how they form such models and, therefore, impact the interaction between humans and computers.

More importantly, cognitive biases can cause harm and open the door to manipulation. Misinformation triggers confirmation bias in Internet users (tendency to seek information that only aligns with one's own beliefs), which lets them believe and propagate such information [60, 164]. Dark patterns [132] and social engineering [23] exploit people's cognitive biases and steer their decision-making. The issue of cognitive biases, hence, becomes a crucial research agenda in HCI to design systems that not only take cognitive biases into account but also remediate their adverse effects [5, 18, 46, 47, 125, 200, 213].

Due to HCI's multidisciplinary nature, research about cognitive biases in HCI is scattered and targets the issue from different angles, methodologies, and application areas. However, there exists no comprehensive review of cognitive bias studies in HCI. In this paper, we provide a scoping review of 127 articles that study cognitive biases in HCI. Our goal is to form a systematic understanding of how cognitive biases manifest in the interaction with computers. Therefore, we analyse the literature and chart how the HCI community conducts research around cognitive biases by categorising papers based on their study focus and application context.

Our results show that HCI research considers cognitive biases as a human factor. HCI researchers aim to understand how human-computer interactions reinforce cognitive biases to inform design considerations that address these biases. We found five different ways HCI researchers engage with cognitive biases (Investigating Effects, Mitigating, Observing, Utilising, and Quantifying) and mapped them in an overarching picture of how HCI researchers study cognitive biases (Figure 1). Computing systems can trigger cognitive biases, which influence, and sometimes steer, real-world human behaviours. Motivated by these concerns, HCI researchers develop **(1) tools and methods** that quantify and capture the occurrences of cognitive biases. These tools and methods help researchers to closer investigate **(2) their effects** on the interaction between humans and computers. With a better understanding of cognitive biases, HCI researchers design **(3) systems and interfaces** that mitigate and leverage cognitive biases, ultimately addressing **(4) real-world human behaviours**. Additionally, we found that cognitive biases are a double-edged sword: not only can their effects steer human judgements, but they are also leveraged for the greater good.

In sum, this paper provides the following contributions:

- We provide a *scoping review* of cognitive bias studies based on a corpus of 127 HCI papers published between January 2010 and May 2024.
- Based on open coding, we derive five ways HCI researchers engage with cognitive biases (Investigating Effects, Mitigating, Observing, Utilising, and Quantifying) and eight application areas where cognitive biases are studied in HCI (Information Interaction and Recommender Systems, Human-AI Interaction, Visualisation, Usability, Behaviour Change, Computer Supported Cooperative Work and Social Computing, Human-Robot Interaction and Autonomous Systems, and Games).
- We map out recommendations and future opportunities for the HCI community to research cognitive biases, voicing the need for community standards, methodological frameworks, and theory-oriented research while discussing the ethical considerations regarding cognitive bias research in HCI. We identify gaps in the literature (Table 2), which guide opportunities for future work.

2 Background

Bias refers to a systematic deviation from the *norm*. There can be several kinds of bias depending on how we set the norm, actors, and application contexts. For example, algorithmic bias describes systematic errors in computing systems that create unfair outcomes [91], and gender bias implies a systematic difference of treatment of one gender over another [41]. In this paper, we focus on *cognitive bias*, first coined by Tversky and Kahneman [188], to refer to a systematic deviation in judgement from the norm of rationality. In this section, we incorporate the literature in cognitive and behavioural science to discuss the notion of cognitive biases, the dual-process theory, and techniques to debias people. We wrap up this section with a discussion of how cognitive biases impact HCI, a review of relevant surveys, and a statement of our contribution to the HCI community.

2.1 Cognitive Biases and Their Interpretations

Humans are not always rational because their cognitive capacity is limited. During the decade of 1950s, Herbert Simon coined the concept of **Bounded Rationality** to explain that, given the complexity of the world and constraints on time and cognitive resources, humans apply **Mental Shortcuts** to faster sift through information and make judgments [166]. Such mental shortcuts can lead to flawed and suboptimal decision-making. Two decades later, Tversky and Kahneman extended this concept and proposed the notion of **Cognitive Bias**: humans employ **Heuristics** as mental shortcuts which systemically deviate their behaviour and the decision-making outcome from the *norm of rationality* [188]. Guaranteeing a fast but suboptimal outcome, heuristics are rules of thumb that humans have adopted through basic instinct, preexisting beliefs, and prior experiences [87]. Through a series of empirical experiments [95, 98, 187–189], Tversky and Kahneman showed that humans employ heuristics and, thus, exhibit several kinds of cognitive biases which systematically skew their decision-making. For example, the anchoring bias makes people rely heavily on the first information presented to them [188]; and the framing effect causes individuals to react to a piece of information differently depending on how it is presented [189]. Subsequent works in cognitive psychology and behavioural science discovered more variations of cognitive biases, such as confirmation bias [142], the halo effect [12] (tendency to rate attractive individuals more favourably for their characteristics), the fundamental attribution error [160] (tendency to overattribute the behaviour of others based on their characteristics), or the Dunning-Kruger effect [113] (tendency to overestimate one's ability despite lacking competence). Up to date, there have been over 180 documented forms[1] of cognitive bias [58]. Meanwhile, recent research has argued that most cognitive biases can be simplified to a form of confirmation bias [147].

Being the by-product of mental shortcuts in decision-making, Kahneman and Tversky viewed cognitive biases as erroneous responses or mental fallacies resulting from the deviation from the norm of rationality [54, 64, 188]. However, recent discourses in psychology have started to shift away from the original interpretation of cognitive biases. The prominent psychologist Gerd Gigerenzer has been a critic of the idea that humans are *biased*, disputing that the norm of rationality does not always exist [194]. In his later works, he argued that heuristics are *fast and frugal* reasoning that helps people make a fast and rational decision at the same time [62, 63]. Some psychologists define cognitive biases as the behavioural consequence of the unconscious, unintended use of mental shortcuts [149, 206]. Hilbert [86] proposes that cognitive biases can be statistically modelled as a result of humans' noisy memory and information processing. In evolution psychology, Haselton and colleagues [78, 79] discuss cognitive biases as rather *design features* of the human mind, citing that humans develop many cognitive biases and heuristics as part of their survival and the natural selection. In this paper, we refer to the more inclusive definition: we consider cognitive biases an **inherent human factor** that broadly and systematically affects and distorts their behaviour.

[1]Not all forms of cognitive bias have been empirically validated in peer-reviewed studies.

2.2 Dual-Process Theory and Debiasing

Human cognition employs two systems of information processing: a fast, automatic, and error-prone **System 1 Thinking** and a slow, conscious, and deliberative **System 2 Thinking** [172, 201]. The so-called **Dual-Process Theory** explains that most of the time, humans resort to using System 1 thinking to make judgments. Psychologists argue that cognitive biases largely emerge from the activation of System 1 thinking [31, 56, 95, 171]. People, therefore, employ heuristics and cognitive biases when sifting through complex information without explicit awareness.

Research in psychology and behavioural science suggests that cognitive biases could be reduced or avoided if people bypass System 1 and shift to System 2 thinking [124, 168, 172]. This can be done through **Cognitive Aid** to guide people to make alternative, rational decisions. Kozyreva et al. [112] propose three main approaches to intervene users away from cognitive biases: nudging, technocognition, and boosting. Nudging [181] changes the environment (i.e., the user interface) and shifts people's behaviour in a subtle way. Notably, nudges can substitute people's autonomous choices with preselected rational decisions. Technocognition [119] are psychological interventions that safeguard people from their biases. For example, slowing down decision-making invites people to reflect on their judgment [170]. Boosting [85] fosters users' metacognition and critical thinking skills to empower control over their decision-making. This also includes education and psychological innoculation [40] that build people's resilience to fast and error-prone thinking.

Correcting people's cognitive biases is, however, not straightforward. Lilienfeld et al. [124] suggest a number of factors that are potential barriers to debiasing. For example, individual differences in working memory and intelligence influence the motivation to engage in rather System 1 or System 2 thinking and, therefore, an individual's receptibility to debiasing interventions [55, 171]. In addition, interventions may not work in the long term and over different contexts [205]. There is also a possibility that interventions may backfire and rather exacerbate the user's existing cognitive biases [207].

2.3 Impact of Cognitive Biases on HCI

Cognitive biases profoundly affect user behaviours, especially when they come into contact with computing systems. The role of cognitive biases influencing the interaction has been discussed in the HCI community over the recent decade [5, 6, 18, 20, 125, 213]. As a result, we observe a growing number of HCI studies investigating cognitive biases (Figure 3). HCI research generally studies the practical aspects of cognitive biases to optimise the human-computer interaction, e.g., how does anchoring bias affect people when they use AI to make decisions [76, 143] or how could nudging interfaces mitigate confirmation bias when people search information on the web [123, 158, 159]. Additionally, interface designers have employed nudges [181], which harness cognitive biases, to steer the user behaviour [28, 109]. Dark patterns [132, 133] and social engineering [23] are interesting case studies where people's cognitive biases are exploited to manipulate their decision-making. Recent works have discussed the notion of *Bias-Awareness* [17, 126] as computing systems take users' existing cognitive biases into account, mitigate their drawbacks, and maximise their benefits.

2.4 Related Surveys

Our review builds upon the existing surveys in the HCI community that ground on cognitive biases, particularly in the areas of behaviour change technologies and human-centred AI. Hekler et al. [83] survey studies on behaviour change published at CHI between 2002-2012. The authors recommend that insights into cognitive biases in behavioural science can inform the design and evaluation of technologies that help people change their behaviours (e.g., eating more healthy food). More importantly, they suggest that HCI and behavioural science literature remain largely siloed within the two communities. On the other hand, the domain of HCI is in a unique position to contribute to the field of behavioural science. Not only can HCI leverage insights in behavioural science, such as cognitive biases, but it can also complement them through ubiquitous sensing, fast prototyping, and data-driven practice. In the same vain, Pinder et al. [152] provide a critical review of digital health behaviour change interventions and suggest that cognitive biases can be leveraged to induce change in health behaviours (e.g., to quit smoking or reduce anxiety) through the use of Cognitive Bias Modification (CBM), which modifies people's subconscious mental shortcuts by gradually modifying attentional bias (tendency to prioritise attention on a certain type of stimuli), approach bias (tendency to approach rather than avoid repetitive cues), and the priming effect (an individual's exposure to one stimulus influences how they respond to a subsequent stimulus). Caraban et al. [28] review HCI studies that proposed and employed technology-mediated nudges to induce behaviour change. They cover 23 mechanisms of nudging by tapping into people's cognitive biases. While nudges are often criticised as manipulating people's autonomy [156, 176], they find that the majority of the nudges rather transparently promote reflective thinking and influence the user choice rather than implicitly manipulating user behaviour.

Researchers have also explored cognitive biases in AI-assisted decision-making [14, 69, 104, 200]. Wang et al. [200] conduct a literature review of concepts in explainable AI (XAI) and synthesise a conceptual framework of how human cognitive patterns drive the need for building XAI and how XAI can alleviate common cognitive biases. Kliegr et al. [104] review and analyse the effects of cognitive biases on the interpretation of machine learning models. Their work suggests that individual differences (e.g., personality traits and numerical literacy) could influence the effectiveness of cognitive bias mitigation. They also highlight that a study of the effects of cognitive biases should precede bias mitigation to ensure that the bias occurs and the bias mitigation strategy does not backfire. Bertrand et al. [14] survey studies that involve cognitive biases in AI-assisted decision-making. The authors provide an overview of the context in which different cognitive biases affect how XAI systems are designed and evaluated in user studies. The work also outlines that XAI systems can mitigate, as well as cause, trigger, or amplify, the users' existing cognitive biases. The authors argue that not all cognitive biases are harmful. Some of them are inherent to the interaction with the AI explanation.

While the prevailing surveys have explored how cognitive biases manifest in human-computer interaction from different angles of HCI, there is a lack of a comprehensive review of cognitive biases throughout the field. Therefore, we propose a scoping review that charts how the broader HCI community studies cognitive biases. In this regard, we review research papers that investigate cognitive biases in different application domains of HCI and draw a framework (Figure 1) of where cognitive biases are situated in the dynamics of human-computer interaction. To the best of our knowledge, we are the first to comprehensively review the research on cognitive biases across different HCI contexts.

2.5 Our Contributions

Limited work has reviewed the issue of cognitive biases in HCI. While this issue is clearly emerging, research has scattered around different angles and application domains due to the multidisciplinary nature of HCI. Therefore, it is essential to summarise these research spans and assemble the big picture of cognitive biases' prevalence in human-computer interaction. In this paper, we present a *systematic* scoping review that highlights the issue of cognitive biases in HCI. Importantly, **our work provides a holistic overview of cognitive biases in the interaction with computer systems.** We differentiate our work from existing surveys in HCI, which address the question in a specific domain and application scenario (e.g., behaviour change or human-AI interaction). To this end, we augment our discussion with insights from the existing surveys and discussions around cognitive biases in HCI.

Our scoping review seeks to understand the question: **how do HCI researchers study cognitive biases?** Through analysing cognitive bias studies throughout different spaces in HCI, we (1) chart the landscape of cognitive biases in HCI, i.e., what aspects of them are studied and leveraged. This helps us to (2) form guidelines on what HCI researchers should consider when researching cognitive biases. Specifically, our work reflects the practice of how HCI researchers have studied cognitive biases. Furthermore, our work identifies (3) challenges and opportunities for HCI research to develop tools, understanding, and designs that take cognitive biases into account to address concerns about human real-world behaviours. We publish our study corpus and the coding manual as supplementary materials for future work to expand upon.

3 Methodology

Our work qualifies as a scoping review [138]. To address our research question, we systematically examine how cognitive bias research is conducted, identify areas or gaps of research, and clarify the notion of cognitive bias in the HCI literature. To conduct this scoping review, one researcher performed (1) database searches, (2) article screening, and (3) data extraction and coding. Three other researchers iteratively cross-checked the process.

3.1 Database Searches

We followed the PRISMA 2020 guidelines [137] to select relevant publications for this scoping review. First of all, we identified HCI research articles published in leading venues in HCI that are likely to publish work on cognitive biases. This includes venues sponsored by ACM SIGCHI (*e.g.,* CHI, CSCW, TOCHI, CHIIR, or IUI), TVCG, IJHCS, and IJHCI. We first started with the search query "*cognitive bias*" in publication titles and abstracts to identify research articles relevant to cognitive biases. We then iteratively derived synonyms based on the initial search. Therefore, we identified new terms: "*human bias*", "*confirmation bias*", and "*decision bias*". We note that the terms "human bias" and "decision bias" are relevant to biases related to humans. To maximise coverage of cognitive bias studies, we included "confirmation bias" as a keyword because of its prevalence as the most common form of cognitive bias. Research in psychology suggests that most cognitive biases can be simplified to confirmation bias [147]. Due to the large number of cognitive biases identified in the literature, we found that the terminologies may differ between different articles. More importantly, some papers do not explicitly use the abovementioned terms in the abstract. By title or abstract, we found that searching these terms returned limited results (only 31 results are returned in the ACM Digital Library, as of 31 May 2024). Therefore, we performed **full-text searches** to get better coverage of articles. However, we did not include the generic term "*bias*" as it returned a large number of irrelevant records from the online databases. After fine-tuning, we performed library searches using the following research query:

```
[[Full Text: "cognitive bias" OR "human bias" OR
"decision bias" OR "confirmation bias"] AND [E-Publication
Date: (01/01/2010 TO 31/05/2024)]]
```

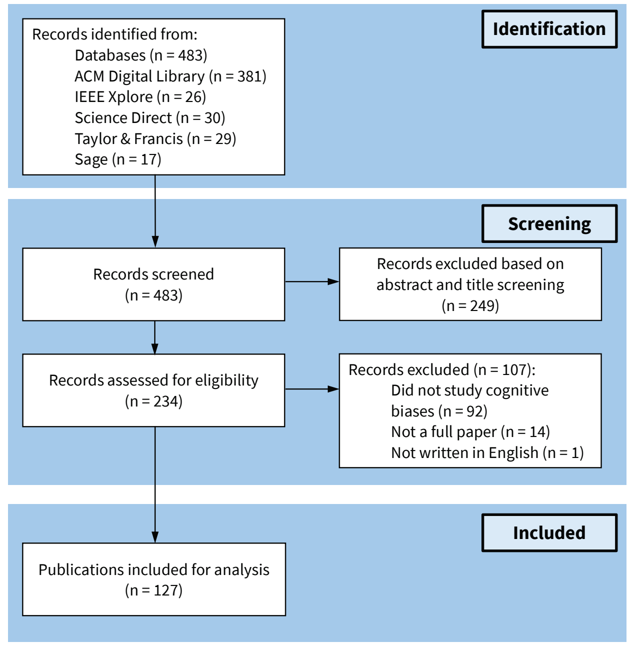

Figure 2: PRISMA 2020 [137] flow diagram for the article screening and selection process.

3.2 Article Screening

We obtained a total of 483 unique records through the database keyword search up to 31 May 2024. We then performed title and abstract screening. At this stage, we included articles that mention keywords that relate to the themes of cognitive biases (including bounded rationality, heuristics, dual-system thinking, decision-making, and mental models). We excluded articles that were clearly outside our scope (e.g., those examining algorithmic bias or media bias). This process left us with 234 papers. Subsequently, we assessed the articles at the full-text level to filter out irrelevant articles that did not investigate cognitive biases. To do so, we excluded papers that did not have cognitive biases as their main focus or study variables (92 papers). In addition, we excluded short papers, posters, late-breaking works, and extended abstracts (14 papers) because they have a different level of maturity than full papers. One paper was excluded because it was not written in English. We derived the final corpus of 127 papers for data extraction. Figure 2 shows a flow diagram for our article selection process.

3.3 Data Extraction and Coding

With our corpus, we created a data extraction sheet to systematically explain our papers from different angles. For each study, we extracted (1) what cognitive biases are studied with their definitions, (2) how cognitive biases are studied, and (3) what application context it covers. We describe the data extraction and coding methodology in the following.

(1) **Cognitive Biases Studied and Their Definitions**: In each paper, we extracted all cognitive biases mentioned with the terms and definitions mentioned by the author(s). We performed a keyword search in the full-text paper to identify possible mentions of cognitive biases. By doing so, we included keywords "bias" and "effect" and cross-checked with Cognitive Bias Foundation's taxonomy of cognitive biases [58], which provides a community-sourced extensive document of more than 180 cognitive biases.

(2) **Study Focus**: For each paper, we extracted how cognitive biases are studied. We then performed open coding and derived five different study focuses in the following bullet points. We note that 14 papers (11.02%) have two study focuses as they consider cognitive biases in multiple angles. For complete information, we publish the data extraction sheet in the supplementary materials.
- **Quantification**: tools, methods, metrics, or mathematical/statistical models to detect, measure, or quantify cognitive biases;
- **Mitigation**: mitigation or prevention of cognitive biases and their adverse effects;
- **Utilisation**: application and utilisation of the effects of cognitive biases in the interaction with computers;
- **Effect Study**: investigation or demonstration of the empirical effects of cognitive biases on the interaction with computers;
- **Observation**: observations or case studies of cognitive biases in people, systems, and their interactions.

(3) **Application Contexts** We extracted the primary application context each paper worked on. Subsequently, we applied open coding to group each paper into eight broad, distinct themes, which represent different areas of HCI research:

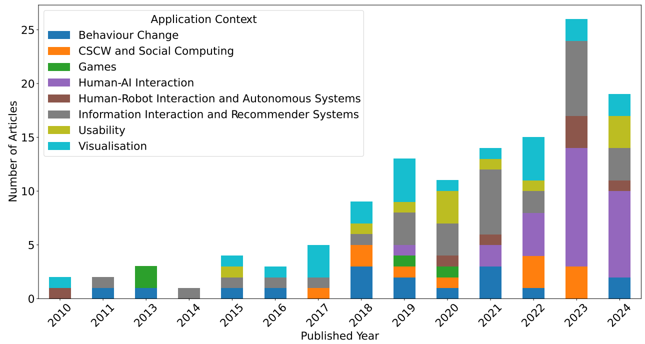

Figure 3: The number of cognitive bias papers by application context and published year, dated from 2010 to May 2024.

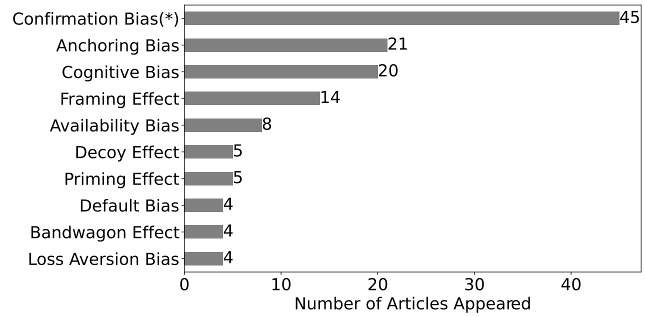

Figure 4: This chart visualises the 10 most frequently mentioned forms of cognitive bias with the number of articles in each of them. Note that some papers may investigate more than one cognitive bias. *The inclusion of "confirmation bias" in the search query may make its occurrences in our corpus more frequent than usual.

(1) Information Interaction and Recommender Systems, (2) Human-AI Interaction, (3) Visualisation, (4) Usability, (5) Behaviour Change, (6) Computer Support Cooperative Work (CSCW) and Social Computing, (7) Human-Robot Interaction and Autonomous Systems[2], and (8) Games.

4 Results

In the following subsections, we provide the analysis of 127 articles in our corpus based on each of our data extraction criteria: publication venue, term and definition of cognitive bias, study focus, and application context.

4.1 Publication Venue and Year

In our corpus, papers published at CHI form the majority of works (40 articles, 31.49%). Other than that, there are papers published at CSCW (12 articles, 9.44%), IJHCI (8 articles, 6.30%), IUI (8 articles, 6.30%), TOCHI (7 articles, 5.51%), CHIIR (6 articles, 4.72%), TVCG (6 articles, 4.72%), and others (40 articles, 31.49%). We observe an upward trend of articles published by year (Figure 3). This reflects that the issue of cognitive biases has increasingly gained attention in HCI research. Notably, roughly half of our corpus (60 articles, 47.24%) was published between 2022 and 2024, with CHI papers making up the majority (27 articles).

4.2 Terms and Definitions of Cognitive Biases

We identified 99 different terms referring to any form of cognitive bias. After merging synonyms (for example, "anchoring bias" and "anchoring effect" are considered the same cognitive bias), we arrived at 92 unique cognitive biases. We found **confirmation bias** to be the most frequently studied (45 articles), which is partially due to its inclusion in the search query. We identified several other cognitive biases, such as anchoring bias (21 articles), the framing effect (14 articles), and availability bias (8 articles). Several papers, however, do not mention any specific form of cognitive biases; for example, 20 articles mention "cognitive bias" or its equivalent terms. Figure 4 charts 10 of the most frequently mentioned cognitive biases in our corpus. We provide the full list of unique cognitive biases in our bias codebook as part of the supplementary materials.

We found discrepancies in the usage of terms of cognitive biases. First of all, *bias*, *effect*, and *heuristics* are used interchangeably. Prominent examples include "anchoring bias" and "anchoring effect" or "availability bias" and "availability heuristics". Goffart et al. [65] cross-termed between "default option effect", "default option bias", and "default bias". We identified many semantically-equivalent terms for "cognitive bias"; for example, "decision bias", "decision heuristics", and "bias in decision making". We also found some cognitive biases to be closely related. There are cross-usages of terms in confirmation bias studies. For example, Liao et al. [123] study *selective exposure* (tendency to focus and seek out information that confirms one's beliefs), which overlaps with confirmation bias. Similarly, Aicher et al. [2] touched on *self-imposed filter bubbles*, a related phenomenon with confirmation bias. Some works also coin context-specific cognitive biases based on existing cognitive biases; for instance, Pafla et al. [151] propose *explanation confirmation bias* based on confirmation bias in explanations provided by AI. We present common cognitive bias synonyms in Table 1.

We discovered some discrepancies in definitions of cognitive bias. Two prominent examples are confirmation bias and anchoring bias. While the majority of papers reference confirmation bias by the definition given in the original work of Nickerson [142], some papers refer to later works in psychology [93] or domain-specific definitions, such as Information Retrieval [5] or Communication Science [106]. Similarly, anchoring bias is mostly defined using the definition in the seminal work of Tversky and Kahneman [188]; however, some papers refer to the definitions in previous HCI research (e.g., [141, 200]). Among articles that refer to the seminal works' definitions, we found variability in the wording. Specifically, many papers frame their cognitive bias definitions as context- or domain-specific. For example, Rieger et al. [158] refer to confirmation bias as "*Users tend to select search results that confirm pre-existing beliefs or values and ignore competing possibilities.*". Naiseh et al. [139] define it as "*Humans favour an XAI classification that is consistent in its output with their beliefs and initial hypothesis.*"

[2]Despite Human-Robot Interaction overlaps significantly with Human-AI Interaction, it deserves a separate category because the physical embodiment of autonomous agents [117].

Table 1: Common cognitive bias synonyms

Cognitive Bias	Semantically Equivalent Cognitive Biases
Cognitive Bias	Decision Bias, Decision Heuristics, Bias in Decision Making, Human Bias
Confirmation Bias	Selective Exposure, Self-Imposed Filter Bubbles, Explanation Confirmation Bias
Anchoring Bias	Anchoring Effect
Availability Bias	Availability Heuristics
Decoy Effect	Asymmetric Dominated Choice
Default Bias	Default Option Bias, Default Option Effect,
Forer-Barnum Effect	Forer Effect, Barnum Effect
Positioning Bias	Positional Bias, Positioning Heuristics
Ambiguity Aversion	Ambiguity Effect
Fundamental Attribution Error	Attraction Effect
Bandwagon Effect	Herd Instinct Bias
Automation Bias	Automation Complacency

Furthermore, we identified 22 instances of cognitive biases without a clear definition stated in the respective papers.

4.3 Study Focuses

We categorised papers in our corpus into five study focuses. Subsequently, these focuses reveal three key narratives of cognitive biases in HCI: (A) computing systems can trigger and mitigate cognitive biases; (B) designers capitalise on cognitive biases in users to steer their behaviours; and (C) HCI researchers develop tools and methods to closer observe cognitive biases. In this section, we review the literature in each narrative based on their study focuses.

A. Computing Systems Can Trigger as well as Mitigate Cognitive Biases in People

4.3.1 Investigating the Effects of Biases. We found 38 papers (29.92%) unveiling cognitive biases that people follow when interacting with computing systems. These studies often set up experiments to demonstrate the effects of cognitive biases, with the goal of understanding cognitive biases as a human factor and deriving design recommendations. Most of the works (30 articles) employed quantitative methods and experimental designs [36, 38, 39, 48, 59, 67, 70, 71, 73, 80–82, 103, 107, 108, 110, 114, 129, 143, 154, 159, 162, 163, 169, 177, 178, 183, 184, 186, 208]. For example, Tomé et al. [186] study loss aversion bias (tendency to avoid losses over achieving equivalent gains) in gameplay and derive design considerations for game designers; Nourani et al. [143] explore anchoring bias in explainable AI and find that users tend to make more errors if they are exposed to system strengths (i.e., they are told that an AI system makes accurate predictions) as an anchor; He et al. [82] investigate the Dunning-Kruger effect on human reliance on an AI system and suggest that people who overestimate their ability tend to rely less on AI advice.

Meanwhile, a limited number of papers study the effects of cognitive biases by using qualitative (2 articles) [61, 173] or mixed methods (6 articles) [37, 43, 50, 68, 134, 144]. For instance, Mendez et al. [134] conduct a qualitative study and a follow-up quantitative experiment to investigate the framing effect in student course selection. Chromik et al. [37] carry out a mixed-method study to examine the illusion of explanatory depth (people's tendency to believe they understand a topic better than they actually do) in explainable AI.

4.3.2 Mitigating Biases. We identified 39 papers (30.71%) that seek to mitigate the effects of cognitive biases. Researchers in HCI employ cognitive aid as a strategy to help users reflect and make rational decisions. More specifically, research proposes tools and user interface designs that serve as cognitive aid [42, 51, 67, 100, 105, 111, 135, 174, 193, 199, 200, 213–216]; for example, Zheng et al. [216] propose an intelligent agent to make the discussion among human teachers more objective and reduce errors in decision-making; and Wang et al. [200] present guidelines for designing explainable AI that encourages its users to avoid amplifying their cognitive biases. Studies also explore using pedagogical tools to teach users critical thinking skills, avoiding their cognitive biases [116, 191, 204]. For instance, Whitaker et al. [204] propose Heuristica, a video game that teaches students to recognise and mitigate cognitive biases using a set of immersive scenarios.

Some research suggests that cognitive biases can be mitigated through system feedback that helps users to reflect on their existing biases [52, 126, 140, 145, 165, 198]. For example, Echterhoff et al. [52] propose a machine learning algorithm that identifies anchored decisions made by users and modifies the presentation order of stimuli to minimise anchoring bias. Narechania et al. [140] develop a visual data analytics tool that shows users their interaction history and encourages them to reflect on their unconscious biases.

We also identified several papers proposing *nudging* [181] to shift users away from biased behaviours. Nudges to mitigate cognitive biases come in the forms of indicators and interface designs [123, 125, 158, 159, 192, 209]. For example, Liao et al. [123]

introduce aspect indicators to reduce selective exposure in information seekers. Rieger et al. [158] employ obfuscation to minimise users' interaction with attitude-confirming search results to mitigate confirmation bias.

Recent works have raised concerns about bias mitigation. Bach et al. [8] propose a list of recommendations for how to incorporate bias mitigation strategies into practical AI applications. Bias mitigation could trigger AI aversion among users and backfire. Therefore, one should consider subtle design patterns, apply bias mitigation periodically rather than constantly, and increase the awareness of potential cognitive biases through the user interface. Some research points out that user-related factors and interaction context could impact the effectiveness of bias mitigation [27, 67, 125, 158, 200]. Graells-Garrido et al. [67] argue that there exists no one-size-fits-all approach to combat cognitive biases. In other words, one bias mitigation strategy does not always work in every individual, context, and scenario. Subsequent studies provide supporting empirical evidence. Cao et al. [27] find that user demographics, such as age and familiarity with probability and statistics, could influence user interaction and, subsequently, amplify the effects of cognitive biases. Rieger et al. [158] point out that situation- and user-related factors (e.g., attitude strength, topic interest, and personality traits) could impact the effectiveness of confirmation bias mitigation approaches.

B. Designers of Computing Systems Capitalise on Users' Cognitive Biases to Steer Behaviours

4.3.3 Utilising Biases. 21 articles (16.53%) leverage cognitive biases to nudge users toward certain behavioural outcomes. We identified two main applications of cognitive biases in the literature: (1) changing user behaviours and (2) incorporating cognitive biases in design. First of all, a significant portion of studies leverage the effects of cognitive biases to shift user behaviours in a predictable way. Lee et al. [118] is among the first to investigate how HCI research could leverage behavioural science to design persuasive technologies. The authors showcase the application of the default bias (tendency to accept what is presented), present bias (tendency to settle for a smaller present reward over a bigger award in the future), and decoy effect (tendency to swap one's preference between two options when a third option is presented) in promoting healthy eating choices. Subsequent studies (e.g., [28, 158, 159, 211, 212, 218]) take the approach of *nudging* [181] – by altering the environment, i.e., the user interface, one can trigger the user's cognitive biases and steer them towards a particular decision or behaviour. For example, Zhang et al. [212] propose an interface nudge to encourage people to reflect on their views on political issues. Zavolokina et al. [211] propose ClarifAI, a tool to nudge users towards more critical news consumption. Meanwhile, some papers take advantage of the effects of cognitive biases to induce behaviour change. Ma et al. [128] employ anchoring bias to promote people's trust in AI. Yamamoto and Takehiro [209] use the priming effect to enhance engagement in critical thinking in web searches. Some research investigates CBM [99, 152], which has been commonly used in psychology to modify people's mental shortcuts towards long-term, habitual behaviour change [94]. Pinder et al. [152] suggest that CBM presents a use-case where cognitive biases are leveraged to change people's habits. Kakoschke et al. [99] propose a CBM-based intervention to reprogram associative links between unhealthy food and automatic appetitive responses, making people eat healthier food.

A notable span of works incorporate cognitive biases in the design of computing systems. Loerakker et al. [127] leverage the framing effect in the design of personal informatics to support self-compassion and positive experiences. Mathur et al. [132] discuss how dark patterns on shopping websites exploit people's cognitive biases and deceive them. Burda et al. [23] investigate cognitive mechanisms of social engineering applications, which manipulate people by triggering their cognitive biases. Theocharous et al. [182] provide a critique of personalised recommendation systems and propose how cognitive biases could be taken into account in these systems.

C. HCI Researchers Develop Tools and Methods to Closer Observe Cognitive Biases

4.3.4 Observing Biases. We found 25 articles (19.68%) that consider cognitive biases as a human factor in the interaction with computers. Some research observe the manifestation of cognitive biases [3, 4, 7, 60, 76, 115, 139, 157, 164]. For example, Rho et al. [157] and Mantri et al. [131] demonstrate the framing effect in user comments in forums of publishers. Haque et al. [76] show that law enforcement agents tended to exhibit anchoring bias when interacting with crime maps presented by decision support systems. Some papers discuss the unintended consequences of cognitive biases arising during the interaction [72, 151, 179]. For instance, Pafla et al. [151] point out the risk of saliency maps triggering confirmation bias when interacting with AI explanations. Habib et al. [72] also suggest that confirm-shaming in cookie consent interfaces (i.e., highlighting negative outcomes of not accepting optional cookies) could target users' loss aversion bias. Moreover, some papers discuss their results from the perspective of cognitive biases and decision-making [32, 164, 185]. Shi et al. [164] study the effect of news veracity on cognitive load. They employ cognitive load as a surrogate of System 2 thinking activation, which links to the manifestation of cognitive biases when processing information. Additionally, we identified a number of survey papers documenting cognitive biases in human-computer interaction, such as behaviour change technology [83], visualisation [44], interactive information retrieval [5], and dark patterns [132].

4.3.5 Quantifying Biases. 18 papers (14.17%) propose methods to detect and quantify cognitive biases. This span of research predominantly sets up experiments that induce cognitive biases and measure cognitive biases using different metrics. A number of studies utilise machine learning algorithms to infer cognitive biases through user interaction data [52, 126, 145, 196, 197]; for example, Wall et al. [196] train Markov models to recognise biased behaviours through the user interaction with scatter-plot visualisation; and Echterhoff et al. [52] use a combination of Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) Neural Networks to capture anchoring bias from sequential decision data. Some papers derived statistical and mathematical modelling as measures of cognitive biases [2, 49, 114, 125, 155]. For instance, Rastogi et al. [155] employ Bayesian modelling for human decision-making in human-AI interaction. In their paper, anchoring and confirmation biases are modelled as scenarios when certain model weights are high. We also found some papers proposing metrics that are derived directly from the original definition in behavioural science or prior literature [25, 33, 45, 88, 129, 186]. Ma et al. [129], for example, quantify overconfidence bias (tendency to have more confidence in one's own abilities) as the difference between the user's expected accuracy of the model and their self-reported self-confidence.

Furthermore, we identified studies that used sensor data to detect cognitive biases [17, 77]. Harris [77] evaluate the bandwagon effect in relevance judgment by using eye-tracking data. Boonprakong et al. [17] employ Functional near-infrared spectroscopy (fNIRS) and feature engineering to measure the effects of cognitive biases when comprehending different opinions.

4.4 Application Contexts

Based on open coding, we identified eight application contexts that span the research space of our corpus. We briefly review the literature in each application context in the following.

4.4.1 Information Interaction and Recommender Systems. We found the majority of cognitive bias studies to be concerned with the area of Information Interaction and Recommender Systems (30 articles, 23.62%). Interestingly, most studies are around the phenomenon of biased information seeking in people, which includes selective exposure [15, 123, 163, 174], misinformation [60, 100, 164], echo chambers [49], and filter bubbles [2]. Some research investigates user attitudes and viewpoints, as people employ them as a principal heuristic for processing information [16, 17, 45, 50, 51]. A significant portion of papers also work on the issue of recommendation systems as they could amplify and have the potential to exploit cognitive biases in users [67, 162, 177, 182]. For example, Graells-Garrido et al. [67] suggest that algorithms and user interfaces should be used in a combination that helps users avoid cognitive mechanisms that lead to biased behaviours.

4.4.2 Human-AI Interaction. The second most popular area is Human-AI Interaction (26 articles, 20.47%). It is a recent and rapidly growing area of study, with most articles published in 2023 and 2024, as shown in figure 3. These papers discuss the problems of explainable AI [37, 71, 73, 143, 151, 200], AI-assisted decision making [8, 25, 27, 35, 52, 128, 129, 155, 178], and trust-reliance in AI [81, 82, 103, 185]. These papers unveiled different cognitive biases that shape the user's mental model when interacting with AI, such as confirmation bias [27, 71, 151, 200], anchoring bias [8, 128, 143, 155], framing effect [52, 73, 103], or the Dunning-Kruger effect [81]. The literature mentions that not only AI systems could trigger and amplify existing cognitive biases in people [8], but other factors, such as limited time [155, 167] and technological expertise [185], could as well influence and facilitate cognitive biases.

4.4.3 Visualisation. We found 22 articles (17.32%) discussing different aspects of cognitive biases in information visualisation. First of all, a number of articles suggest that cognitive biases impact how users interact with and make decisions based on the visualised information [36, 43, 107, 108, 134, 154, 192]. For example, Kong et al. [108] investigate the effects of confirmation bias on how people recall the visualisation of titles. Mendez et al. [134] suggest that the framing effect in visualisation can induce students to put more effort into course selection. Notably, some research proposes that cognitive biases can be detected and quantified through the user's behaviours [145, 196, 197]. For instance, a series of works by Wall et al. [196, 197] propose computational methods (e.g., Markov models) to characterise and predict cognitive biases from the systematic deviation in user interaction from a theoretical baseline. Studies also discuss how cognitive biases can inform the design of interactive visualisation systems [9, 197, 199] as well as the presentation of information [43, 108, 192, 213] to avoid unintended effects from visualisation. Moreover, some research discusses the symbiosis of cognitive biases and visualisation systems, as we can use either of them to improve the other [140, 145, 197].

4.4.4 Behaviour Change. We identified 16 articles (12.60%) that focused on behaviour change. These papers highly overlap with papers that utilise cognitive biases (12 articles, see Table 2), as they seek to shift user behaviour by tapping into users' cognitive biases. Notably, works by Lee et al. [118] and Hekler et al. [83] pioneer how cognitive biases can be used to induce behaviour change in HCI, for example, to encourage healthy habits, Lee et al. [118] design a webpage for snack buying that shows two healthy food choices on the first page, requiring users to click next to see other food options. This taps into users' default bias and steers their food selection behaviour. Later research discusses cognitive biases in nudges and persuasive technologies [28, 29, 65, 109, 156, 158, 212, 218]. Some research expands the discussion on CBM to induce long-term behaviour change in health-related behaviours [99, 152]. Moreover, some papers discuss the potential of cognitive assistants in boosting reasoning and critical thinking skills in people [116, 211].

4.4.5 Usability. Eleven articles (8.66%) investigate cognitive biases from the angle of usability. We found a significant portion of research discussing how cognitive biases can influence the usability of interactive systems. For example, Veytizou et al. [193] suggest that the halo effect can influence user opinions on usability. A series of works by Mathur and colleagues [132, 133] discuss several cognitive biases (e.g., anchoring bias, bandwagon effect, or default bias) that could be exploited by dark pattern user interfaces. Alqahtani et al. [4] study uncertainties in the interaction with self-tracking systems. They argue that users may rely on heuristics and cognitive biases (e.g., confirmation and availability bias) as strategies to avoid uncertainties in the interaction. Chen et al. [33] investigate the Weber-Fechner law as a cognitive bias that influences the perceived visual consistency when users view visual icons across different devices caused by adaptive scaling.

4.4.6 Computer-Supported Cooperative Work (CSCW) and Social Computing. Eleven articles (8.66%) focus on the interactions beyond an individual. We found a notable portion of papers discuss the impact of cognitive biases in crowdsourcing and collective ratings [3, 34, 75, 88, 184], suggesting that humans (i.e., annotators) have a potential to introduce biases into data and algorithms. For example, Hube et al. [88] and Thomas et al. [184] suggest that different cognitive biases can impact and introduce errors to data annotations. Haq et al. [75] propose a method to mitigate errors from cognitive biases in data workers. In addition, some papers investigate cognitive biases in the context of human collaborative technologies. For instance, Shi et al. [165] suggest that activity

Table 2: Categorisation of cognitive bias papers by study focus and application context with their respective count and references. Note that some papers have more than one study focus, therefore, they can have multiple entries in a row. Meanwhile, papers categorised by application context are mutually exclusive. (n/a means no paper examining in that category)

Study Focus × **Application Context**	**Mitigation** (N=39)	**Effect Study** (N=38)	**Observation** (N=25)	**Utilisation** (N=21)	**Quantification** (N=18)
Information Interaction and Recommender Systems (N=30)	10 articles [51, 67, 100, 123, 125, 153, 158, 174, 209, 214]	10 articles [50, 61, 67, 70, 144, 162, 163, 169, 177, 208]	4 articles [5, 60, 164, 179]	3 articles [15, 182, 209]	6 articles [2, 17, 45, 49, 77, 125]
Human-AI Interaction (N=26)	7 articles [8, 27, 35, 52, 155, 190, 200]	11 articles [37, 71, 73, 81, 82, 103, 110, 129, 143, 178, 183]	6 articles [26, 76, 139, 151, 167, 185]	1 article [128]	4 articles [25, 52, 129, 155]
Visualisation (N=22)	10 articles [11, 42, 111, 126, 140, 145, 192, 198, 199, 213]	6 articles [36, 43, 107, 108, 134, 154]	3 articles [9, 44, 131]	3 articles [127, 134, 192]	4 articles [126, 145, 196, 197]
Behaviour Change (N=16)	2 articles [116, 159]	2 articles [68, 159]	1 article [83]	12 articles [28, 29, 65, 99, 109, 118, 146, 152, 156, 211, 212, 218]	n/a
Usability (N=11)	3 articles [105, 193, 215]	n/a	5 articles [4, 32, 72, 121, 133]	2 articles [23, 132]	1 article [33]
CSCW and Social Computing (N=11)	5 articles [75, 88, 135, 165, 216]	3 articles [59, 173, 184]	3 articles [3, 115, 157]	n/a	1 article [88]
Human-Robot Interaction (N=7)	n/a	4 articles [38, 48, 80, 114]	3 articles [7, 84, 150]	n/a	1 article [114]
Games (N=4)	2 articles [191, 204]	2 articles [39, 186]	n/a	n/a	1 article [186]

traces can help mitigate cognitive biases in peer evaluations. Zheng et al. [216] point out that incorporating AI in group decision-making can stimulate human members to reflect on their logic and reduce cognitive biases in flawed decision-making.

4.4.7 Human-Robot Interaction and Autonomous Systems. Seven articles (5.51%) discuss cognitive biases in the interaction with robots and/or physical autonomous systems [7, 38, 48, 80, 84, 114, 150]. For example, Paepcke and Takayama [150] find that confirmation bias affects how users set expectations about the robot's ability. Hayashi et al. [80] show that anchoring bias makes people stick with human experts' suggestions for decision-making over those suggested by robots. Some research also investigates cognitive biases when acting with autonomous vehicles [38, 48]. Colley et al. [38] suggest that the presence of autonomous vehicles could trigger the halo effect in pedestrians who signal with the vehicle. Interestingly, we found no work that seeks to mitigate cognitive biases in the domain of human-robot interaction.

4.4.8 Games. Four articles (3.14%) discuss the impact of cognitive biases in gameplay. Constant and Levieux [39] find that dynamic game difficulty adjustment could trigger players' overconfidence bias and illusion of control. Tomé et al. [186] study how lost aversion bias impacts how players make decisions in games. Interestingly, some research suggests games could be incorporated into learning systems to mitigate cognitive biases [191, 204]. Veinott et al. [191] examine how serious video games can improve people's ability to be aware of and, therefore, overcome their own cognitive biases.

5 Discussion

In this section, we reflect on higher-level insights obtained from our analysis of the articles in our corpus. We address the main research question of how HCI researchers study cognitive biases, explain the role of cognitive biases as a double-edged sword in the interaction with computers, and discuss the ethical implications of the exploitation of cognitive biases in users of computing systems.

5.1 How Do HCI Researchers Study Cognitive Biases?

The literature clearly indicates that cognitive biases are pervasive in HCI. Humans are not always rational. They are susceptible to making flawed decisions. When interacting with computing systems, user interfaces have the potential to trigger users' cognitive biases. Subsequently, cognitive biases systematically affect users' mental models and, subsequently, their real-world behaviours when interacting with computing systems. According to our findings in section 4.3, the manifestation of cognitive biases in HCI can be explained in three layers: **(A) systems could trigger or remedy existing cognitive biases in users**; **(B) interface designers capitalise users' cognitive biases to (either intentionally or not) steer and manipulate their behaviours**; and **(C) HCI researchers observe cognitive biases in the interaction with computers and develop tools and methods to closer study them.** Although these scenarios do not always happen simultaneously, they all highlight that cognitive biases are a crucial human factor in designing computing systems and user interfaces.

HCI researchers study cognitive biases in the interaction between humans and computers, not only to understand them as a human factor, but also to inform the design of computing systems to better adapt to the user's mental models. By deriving insights from behavioural economics and psychology, HCI research develops tools and metrics to detect, quantify, and study cognitive biases in human-computer interaction more closely. Moreover, many HCI papers employ cognitive biases as a tool to study human behaviours and decision-making. In line with the recent literature in psychology, HCI researchers treat cognitive biases as features of the human mind [78, 79] and, therefore, incorporate them as a human factor. In sum, insights from cognitive bias studies help the HCI community derive recommendations and practicality for designs that take biases in people into account. Recent research [17, 126, 217] has introduced the notion of *bias-awareness*, which refers to the ability to detect, understand, and take into account cognitive biases in people and computing systems. HCI researchers leverage bias-aware systems to address human behaviour and its associated real-world concerns, such as helping humans make objective decisions [155, 165], calibrating their trust in AI [103, 185], or guiding them how to discern online propaganda [20, 211]. Figure 1 summarises how HCI researchers work with cognitive biases to (1) develop tools and methods, (2) better understand people and their biases, (3) inform the design of computing systems and user interfaces, and (4) address real-world human behaviours.

5.2 Cognitive Biases as a Double-edged Sword in HCI

The prevalence of cognitive biases in HCI is a double-edged sword; there are negative and positive effects arising from cognitive biases. A significant number of articles in our corpus outline how cognitive biases could result in negative consequences, such as undermining the collaboration between human and AI systems [81, 82, 103, 185, 200], facilitating the spread of misinformation [60, 100, 164] and unhealthy information behaviours [15, 123, 163, 174], inducing errors when navigating through information [36, 43, 107, 108, 134, 154, 192], or affecting the quality of crowdsourced data [3, 34, 75, 88, 184]. Subsequently, our review identifies different methods proposed to mitigate the adverse effects of cognitive biases. HCI researchers study and employ *cognitive aid*, as introduced in psychology, to help users reflect on themselves and make informed decisions. Based on the dual-system theory, these systems shift people towards using the slower and more deliberative System 2 thinking rather than the fast and error-prone System 1 thinking to make decisions. We found such cognitive aid comes in the form of either *nudging* or *boosting*. While nudging guides people to shift their behaviour, boosting empowers their cognitive and motivational compentencies [85, 112]. The latter approach, boosting, appears in our review as, for instance, tools to teach users critical thinking skills and how to spot their cognitive biases [116, 191, 204].

On the other hand, cognitive biases can benefit the interaction. Our review identifies different studies and research leveraging cognitive biases for the greater good. One prominent example, which is mentioned above, is nudging. Nudging capitalises on the users' cognitive biases to steer them towards a certain behavioural outcome [28]. At the same time, nudges present a use-case where **different cognitive biases can cancel each other out**. Rieger et al. [158] employ nudges as targeted obfuscation in search results to decrease user interaction with attitude-confirming information. Explained by [28], this nudge triggers status-quo bias, which prevents users from interacting with the obfuscated items, and, therefore, mitigates confirmation bias. Furthermore, cognitive biases can be leveraged to induce long-term behaviour change in the form of CBM [99, 152], which has been largely used in healthcare and intervention-focused approaches (e.g., helping individuals quit smoking, eat healthier, or alleviate anxiety symptoms).

Our findings, therefore, suggest that humans are cognitive misers. Some forms of cognitive bias can present in users and their interaction with computing systems. We recommend that the HCI community designs systems aiming to mitigate the negative effects of biases while considering what benefits we can leverage from the users' existing cognitive biases.

5.3 Ethical Considerations from the Exploitation of Cognitive Biases in People

We must, however, acknowledge the ethical implications arising from exploiting inherent human biases. Humans often exhibit cognitive biases without explicit awareness. Therefore, designs and technologies that harness these cognitive biases could risk manipulating their behaviours. Richard Thaler, who first coined the term *nudges*, discussed that the same techniques used to nudge people could be used for negative intentions – the so-called *sludges* [180]. Nudges, on the other hand, could harm user autonomy by steering their behaviours without their awareness and consent [21, 156, 175, 176]. Daniel Kahneman himself and other psychologists also criticised nudges as potential *benevolent paternalism*: governments or ruling institutions can employ nudges to manipulate individuals' choices by assuming the "best interest" of the people [92, 95, 176]. Dark patterns [132] and social engineering [23, 57] are well-researched practical examples where cognitive biases are misused to influence people's decision-making. Boonprakong et al. [17, 19] argue that the same techniques to detect and mitigate cognitive biases could be used to reaffirm and steer people's beliefs. The Cambridge Analytica scandal [13] demonstrate that people's attitudes could be derived from social media interaction data and, in turn, used to target and sway their opinion-making.

Designers and HCI practitioners shape user experiences and build systems that steer user behaviour. Therefore, they should be held accountable for the ethical implications arising from their design choices. Designers should be well aware that users are inherently susceptible to cognitive biases and of what harm they potentially cause [19] (e.g., people who can fall victim to misinformation are less likely to be debunked [120]). One solution to address ethical concerns could be promoting transparency, which gives users the awareness of their cognitive biases being used. Bias-exploiting interfaces can practically ask for informed consent from users that their behaviour may be subconsciously steered. Zhu et al. [217] suggest that, by giving the awareness of how systems collect and process data, users can make informed decisions. Moreover, legal restrictions, such as the European Union's General Data Protection Regulation (GDPR)[3], could also limit how much (sensitive) data systems can collect for bias detection and quantification.

6 Recommendations to the HCI Community

HCI is uniquely at the intersection of multiple disciplines. Our review suggests that the HCI community derives definitions and theories from psychological and behavioural sciences. However, there exists a gap between HCI and these fields. In this section, we discuss the need for the community to establish a standard for bias terminologies and to closer engage with behavioural science and psychology.

6.1 The Use of Cognitive Bias Terminologies

Our review maps out a series of discrepancies in the use of cognitive bias terminologies and definitions. Variations from the standard terminology, including the use of self-defined terms, could cause confusion among the readers and those who search literature using certain keywords. We originally derived only a few records (for example, N = 31 on the ACM Digital Library) when performing a keyword search based on title or abstract. For this reason, we extended our search to records from full-text searches, which returned significantly more results. This implies that (terms for) cognitive biases are varied and often only mentioned in the full-text paper. We suggest that authors in the HCI community should (1) clearly mention the cognitive biases they study, (2) give explicit definitions, and (3) provide a connection to the notion of cognitive biases. By providing clarity and connection to psychology and behavioural science, HCI researchers can improve the internal validity of their studies. Nonetheless, some psychologists argue that the field of psychology itself is experiencing a similar problem, as new constructs are redundantly invented for existing psychological constructs [53, 74].

The lack of a clear definition and reference to cognitive biases also poses a problem. While cognitive biases are relatively well-known phenomena, we recommend authors in the HCI community clearly define cognitive biases and provide definitions and references to make sure that their studies are theoretically grounded. We suggest that, if possible, definitions could link to the seminal works in behavioural economics and psychology; for example, linking anchoring bias with the seminal work of Tversky and Kahneman [188] or confirmation bias with the work of Nickerson [142]. Furthermore, with some cognitive biases being made context-specific (e.g., confirmation bias as "Selective Exposure" or "Self-Imposed Filter Bubbles", and the Dunning-Kruger effect as "Illusion of Explanatory Depth"), we suggest that authors should make clear that they study such cognitive biases in a specific context.

We argue that it is necessary to establish a community standard for terminologies and definitions. For example, what is the distinction between *heuristics* and *cognitive biases*? Should we use "Ambiguity Effect" or "Ambiguity Aversion"? More importantly, our findings show that there is a discrepancy in understanding whether cognitive biases, as a human factor, are heuristics people use to make faster decisions or *consequences* from the use of such heuristics. Most papers say "mitigate cognitive biases": does it mean we mitigate the cognitive bias itself or its effects? As the notion of cognitive biases is increasingly discussed in the HCI community, we envision that the community could find a consensus on the best practices to report research regarding cognitive biases.

6.2 Closer Engagement with Behavioural Science and Psychology

Cognitive biases are grounded in the fields of behavioural, psychological, and cognitive sciences. Therefore, our review voices a need for the HCI community to connect with the literature and scholars in these domains. Prior research suggests that insights in behavioural and cognitive sciences can inform and integrate with the HCI field to conduct studies that are grounded in theory rather than relying on intuition [83, 155, 213]. Because our understanding of human decision-making has been limited, we envision that HCI research can complement behavioural science. With the ability to fast prototyping and running user studies (such as A/B testing), HCI researchers can quickly verify behavioural science theories [83]. The field of HCI also offers multidisciplinary perspectives that augment the traditional understanding of cognitive biases. Tomé et al. [186] discuss that, while loss aversion bias has been studied in behavioural science, we have little understanding of how it affects gameplay.

Recent research in psychology has signalled a shift away from Kahneman and Tversky's original interpretation of cognitive biases [63, 79, 203]. Different schools of psychologists (e.g., Kahneman & Tversky [188] vs. Gigerenzer [62]) may view the issue of how humans satisfy their cognitive constraints differently. The notion of cognitive bias, therefore, may not offer the most robust explanation that fits human behavioural effects/phenomena [161, 202]. While the HCI community has widely adopted the traditional notion of cognitive biases as an explainer of HCI-related effects (e.g., selective exposure, information framing, or dark patterns), the field could also consider and keep up with the more recent or inclusive definitions, such as Gigerenzer's *fast and frugal heuristics* [62], the challenge of humans' cognitive limitations [166], or noise in human decision-making [96].

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

7 Avenues for Future Research

Our review signals multiple paths for future research on cognitive biases in HCI. In this section, we discuss the potential for establishing frameworks to study cognitive biases, considerations of effectively leveraging and mitigating cognitive biases, under-explored application contexts, and improving the external validity of cognitive bias studies in HCI.

7.1 Methodological and Theoretical Framework for Studying Cognitive Biases in HCI

Limited research (18 papers, 14.17% of corpus) explores methods, tools, and frameworks to quantify cognitive biases in the interaction with computers. Our findings show that different papers pursue different approaches to quantifying cognitive biases through metrics, statistical modelling, and physiological sensors. Yet, these methods tend to be catered specifically to particular cognitive biases (e.g., anchoring bias) or application contexts (e.g., interaction with information visualisation or AI-assisted decision-making). We suggest that there could be quantification tools that are agnostic to particular cognitive biases or scenarios. Future research could also explore tools to indicate cognitive biases as simply a *deviation from the norm of rationality*. Similarly, Liu [125] proposes a probabilistic framework for human fairness in decision-making. Boonprakong et al. [17] investigate physiological expressions of *any* cognitive biases in opinion comprehension.

7.2 Considerations of Effectively Leveraging and Mitigating Cognitive Biases in HCI

7.2.1 Leveraging Cognitive Biases. Limited works have explored how cognitive biases can be harnessed for the greater good. Our understanding of cognitive biases in HCI is emerging; therefore, we envision the HCI community could avoid the harm of cognitive biases and leverage their benefits. From the literature, we identify two ways that cognitive biases can be leveraged. Firstly, by understanding how cognitive biases affect human behaviours, HCI researchers can take cognitive biases into account when designing interactive systems. Cognitive biases can be applied to induce systematic behaviour changes for the greater good, such as critical thinking engagements [209, 211, 212], healthy food diet [99], or support self-compassion [127]. Secondly, cognitive biases are used to steer users' behaviour in the form of nudges. Caraban et al. [28] document 23 different mechanisms of nudging and discuss cognitive biases associated with each type of nudge. However, research in behavioural science suggests several shortcomings of nudges. Specifically, the effectiveness of nudges may be limited and not sustained over time [10, 21]. At the same time, a plethora of individual and contextual factors may positively or negatively influence the occurrences of cognitive biases [19, 27, 158] and, subsequently, compromise the success of nudges. We, therefore, argue that future research could (1) conduct longitudinal studies to evaluate the effectiveness of nudges and (2) investigate how we could consider individual and contextual factors in utilising cognitive biases effectively. Ultimately, by leveraging the effects of cognitive biases and avoiding their harm, we could enhance the capability of humans and optimise the symbiosis between humans and machines.

7.2.2 Mitigating Cognitive Biases. Research in HCI has pointed out how cognitive biases can negatively affect the interaction with computers. In response, various studies also propose methods to alleviate these effects. However, our findings suggest many research gaps, which echo the discussions in psychology and behavioural science regarding barriers to effective debiasing [124, 205]. First of all, some bias mitigation techniques, such as providing system feedback or pedagogical tools, have been evaluated in only certain contexts like visualisation and human-AI interaction. Future research could assess the effectiveness of the same set of techniques across different application contexts like information interaction or gameplay. For example, Narechania et al. [140] suggest that pointing users to their interaction history in a visual analytics tool could help users reflect on their existing cognitive biases. However, the question remains whether the same approach works if social media users are presented with browsing history.

Secondly, limited research has investigated whether bias mitigation works in practice. Bach et al. [8] suggest a set of recommendations when incorporating bias mitigation into real-world applications. Recent research in behavioural science has also discussed the danger of the *backfire effect*, which could unexpectedly overturn the effectiveness of an intervention [1, 24]. Future research could investigate when and how the backfire effect occurs when mitigating cognitive biases.

7.3 Understudied Application Contexts

Our findings indicate a number of areas with limited research. Table 2 suggests research fixation and gaps. Based on application context, we found that limited research has investigated the contexts of behaviour change, usability, CSCW, human-robot interaction, and games. Research around behaviour change has predominantly focused on utilising cognitive biases. Meanwhile, limited research seeks to quantify the effects of cognitive biases when steering user behaviour. Most research in our corpus considers one side of the picture – either quantifying either the effects of cognitive biases or the effectiveness of behaviour-change interventions – assuming that cognitive biases take effect regardless of the individual and interaction contexts. With the ability to quantify the effects of cognitive biases, one can empirically measure to what extent people's behaviour has changed and how strong the effect is. Similarly, some papers in our corpus showcase how the angles of bias mitigation and quantification can be harmonised [52, 88, 125, 126, 145, 155]. We envision that future research could consider multiple angles to closer study cognitive biases.

We found limited research in the realm of usability that considers cognitive biases, although this issue is central in HCI research. We suggest more research could explore interface design elements that trigger and reinforce cognitive biases (e.g., [132, 136] discuss how dark patterns are connected with certain cognitive biases). Also, no research in our corpus discusses creativity in conjunction with cognitive biases, such as the issue of design fixation [90] – a cognitive bias that makes people stick to a set of pre-conceived ideas and restrict the choices of design. Some HCI scholars [102, 195] have empirically investigated design fixation, however, they make minimal connection with the discourse around cognitive biases.

Additionally, future research could consider cognitive biases beyond just humans and systems, specifically in the domains of CSCW and human-robot interaction. While some works employ bias mitigation strategies in human collaborative work [165, 216], limited research has explored where these biases come from and the potential to leverage them. Future research may address the question of how do computing systems systematically trigger cognitive biases in a crowd of users (e.g., human teams or social network users), and how cognitive biases can be leveraged in coorperative tasks.

7.4 Expanding External Validity

There are a number of threats to the external validity of existing cognitive bias studies in HCI. First, most studies are conducted in a controlled environment. Outside of the laboratory, a plethora of external factors could affect the way people exhibit cognitive biases. We suggest future research consider running in-the-wild studies to reflect how cognitive biases manifest in real-world interactions. In addition, only a few papers consider multiple forms of cognitive bias in conjunction. Humans could exhibit more than one cognitive bias at the same time (e.g., [213]), while multiple cognitive biases can interact, reinforce, or cancel each other [28, 213]. Future research could conduct studies that consider possible cognitive biases that could occur and confound the study variables. For example, studying confirmation bias in social media browsing might introduce anchoring bias when viewing the contents in a sequence and overconfidence bias when the user has more expertise in the content topic. With the awareness of potential cognitive biases in an experiment, researchers could consider ways to minimise these confounds, such as counterbalancing (anchoring bias) and taking topic expertise as a control variable (overconfidence bias).

8 Limitations

This paper has several limitations. First of all, our corpus does not exclusively cover all cognitive bias studies in HCI. We believe, however, that the inclusion of SIGCHI-sponsored venues (such as CHI, CSCW, IUI, or CHIIR) gives a representative view of the HCI community's discourse around cognitive biases. Moreover, the choice of search keywords may not cover all cognitive biases. The literature refers to cognitive biases in many different ways [14]. It is possible that some papers investigate a relevant issue around cognitive biases but do not explicitly mention the term cognitive biases, for example, bounded rationality [101, 130, 148], decision-making fairness [66, 210], systematic bias [89], design fixation [195], self-selection bias [217], or selective exposure [122]. We share the same sentiment with Kliegr et al. [104], who argue there is an abundance of cognitive phenomena that are not regarded as cognitive biases. We also acknowledge that, since the definition of cognitive biases and heuristics (according to Tversky and Kahneman's original school of thought [188]) has been challenged by many psychologists (e.g., [62, 79, 96]), HCI studies may move away from such concepts and use other terms. While our scoping review mainly investigates the use of cognitive biases in HCI research, we reflect that the actual literature around cognitive biases utilises a diverse range of terms that may not be included in our review.

The list of cognitive biases studied and their figures (Figure 4) may be subject to discussion and change as the research landscape is evolving and some cognitive biases could be considered as a specialised form of another cognitive bias. For example, recency bias (the tendency to more easily remember what happened recently) is considered a form of the peak-end rule. To the best of our knowledge, there has been no commonly agreed-upon taxonomy for cognitive biases, with the taxonomy of the Cognitive Bias Foundation [58] providing extensive coverage of more than 180 cognitive biases. Additionally, we acknowledge that the results could be subject to screening biases and personal views as only one researcher performed article screening and coding.

9 Conclusion

Humans employ heuristics and mental shortcuts to effectively make decisions under their inherent limited cognitive capacity. These shortcuts result in cognitive biases, which systematically influence how humans interact with computers. The HCI community has increasingly discussed this issue in the recent decade. This scoping review charts how HCI researchers study cognitive biases. From 127 articles identified, we found that the prevalence of cognitive biases in HCI gives opportunities for researchers to study, mitigate, and leverage their effects to inform designs, optimise the interaction, and address real-world human behaviour. The literature suggests that cognitive biases are a two-edged sword. While we can leverage their effects to induce behaviour change, the same mechanism can be used to manipulate people's decision-making and harm their autonomy. Our results reveal various terminologies and definitions for cognitive biases, suggesting a lack of standards for terming and defining cognitive biases in HCI. To this end, our findings promise several avenues for future research to better understand cognitive biases in the interaction between humans and computers and the need to connect with the literature in behavioural science and psychology.

Acknowledgments

We thank the anonymous reviewers and members of the University of Melbourne's HCI group for their suggestions, which positively helped shape this scoping review.

References

[1] 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

[2] Annalena Bea Aicher, Daniel Kornmüller, Wolfgang Minker, and Stefan Ultes. 2023. Self-imposed Filter Bubble Model for Argumentative Dialogues. *Proceedings of the 5th International Conference on Conversational User Interfaces* (2023), Article 23. Publisher: Association for Computing Machinery.

[3] 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. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3502040 Publisher: Association for Computing Machinery.

[4] Deemah Alqahtani, Caroline Jay, and Markel Vigo. 2020. The Role of Uncertainty as a Facilitator to Reflection in Self-Tracking. *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (2020), 1807–1818. https://doi.org/10.1145/3357236.3395448 Publisher: Association for Computing Machinery.

[5] Leif Azzopardi. 2021. Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval. *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (2021), 27–37. https://doi.org/10.1145/3406522.3446023 Publisher: Association for Computing Machinery.

[6] Leif Azzopardi and Jiqun Liu. 2024. Search under Uncertainty: Cognitive Biases and Heuristics - Tutorial on Modeling Search Interaction using Behavioral Economics. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval* (Sheffield, United Kingdom) *(CHIIR '24)*. Association for Computing Machinery, New York, NY, USA, 427–430. https://doi.org/10.1145/3627508.3638297

[7] Franziska Babel, Philipp Hock, Katie Winkle, Ilaria Torre, and Tom Ziemke. 2024. The Human Behind the Robot: Rethinking the Low Social Status of Service Robots. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3610978.3640763

[8] Anne Kathrine Petersen Bach, Trine Munch Nørgaard, Jens Christian Brok, and Niels van Berkel. 2023. "If I Had All the Time in the World": Ophthalmologists' Perceptions of Anchoring Bias Mitigation in Clinical AI Support. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3581513 Publisher: Association for Computing Machinery.

[9] Aruna D. Balakrishnan, Susan R. Fussell, Sara Kiesler, and Aniket Kittur. 2010. Pitfalls of Information Access with Visualizations in Remote Collaborative Analysis. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (2010), 411–420. https://doi.org/10.1145/1718918.1718988 Publisher: Association for Computing Machinery.

[10] Oswald Barral, Gabor Aranyi, Sid Kouider, Alan Lindsay, Hielke Prins, Imtiaj Ahmed, Giulio Jacucci, Paolo Negri, Luciano Gamberini, David Pizzi, and Marc Cavazza. 2014. Covert Persuasive Technologies: Bringing Subliminal Cues to Human-Computer Interaction. In *Persuasive Technology*, Anna Spagnolli, Luca Chittaro, and Luciano Gamberini (Eds.). Springer International Publishing, Cham, 1–12.

[11] Eric P. S. Baumer, Jaime Snyder, and Geri K. Gay. 2018. Interpretive Impacts of Text Visualization: Mitigating Political Framing Effects. *ACM Trans. Comput.-Hum. Interact.* 25, 4 (2018). https://doi.org/10.1145/3214353

[12] Neil E. Beckwith and Donald R. Lehmann. 1975. The Importance of Halo Effects in Multi-Attribute Attitude Models. *Journal of Marketing Research* 12, 3 (1975), 265–275. http://www.jstor.org/stable/3151224

[13] H. Berghel. 2018. Malice Domestic: The Cambridge Analytica Dystopia. *Computer* 51, 05 (may 2018), 84–89. https://doi.org/10.1109/MC.2018.2381135

[14] Astrid Bertrand, Rafik Belloum, James R. Eagan, and Winston Maxwell. 2022. How Cognitive Biases Affect XAI-Assisted Decision-Making: A Systematic Review. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (Oxford, United Kingdom) *(AIES '22)*. Association for Computing Machinery, New York, NY, USA, 78–91. https://doi.org/10.1145/3514094.3534164

[15] Md Momen Bhuiyan, Michael Horning, Sang Won Lee, and Tanushree Mitra. 2021. NudgeCred: Supporting News Credibility Assessment on Social Media Through Nudges. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (2021). https://doi.org/10.1145/3479571

[16] Markus Bink, Sebastian Schwarz, Tim Draws, and David Elsweiler. 2023. Investigating the Influence of Featured Snippets on User Attitudes. *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (2023), 211–220. https://doi.org/10.1145/3576840.3578323 Publisher: Association for Computing Machinery.

[17] 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. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3580917 Publisher: Association for Computing Machinery.

[18] Nattapat Boonprakong, Gaole He, Ujwal Gadiraju, Niels Van Berkel, Danding Wang, Si Chen, Jiqun Liu, Benjamin Tag, Jorge Goncalves, and Tilman Dingler. 2023. Workshop on Understanding and Mitigating Cognitive Biases in Human-AI Collaboration. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 512–517. https://doi.org/10.1145/3584931.3611284

[19] Nattapat Boonprakong, Saumya Pareek, Benjamin Tag, Jorge Goncalves, and Tilman Dingler. 2025. Assessing Susceptibility Factors of Confirmation Bias in News Feed Reading. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3706598.3713873

[20] 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

[21] Luc Bovens. 2009. *The Ethics of Nudge.* Springer Netherlands, Dordrecht, 207–219. https://doi.org/10.1007/978-90-481-2593-7_10

[22] Andreas Bulling and Thorsten O. Zander. 2014. Cognition-Aware Computing. *IEEE Pervasive Computing* 13, 3 (2014), 80–83. https://doi.org/10.1109/mprv.2014.42

[23] 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

[24] Sahara Byrne and Philip Solomon Hart. 2009. The Boomerang Effect A Synthesis of Findings and a Preliminary Theoretical Framework. *Annals of the International Communication Association* 33, 1 (2009), 3–37. https://doi.org/10.1080/23808985.2009.11679083 arXiv:https://doi.org/10.1080/23808985.2009.11679083

[25] Federico Cabitza, Andrea Campagner, Riccardo Angius, Chiara Natali, and Carlo Reverberi. 2023. AI Shall Have No Dominion: On How to Measure Technology Dominance in AI-Supported Human Decision-Making. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3581095 Publisher: Association for Computing Machinery.

[26] Ángel Alexander Cabrera, Marco Tulio Ribeiro, Bongshin Lee, Robert Deline, Adam Perer, and Steven M. Drucker. 2023. What Did My AI Learn? How Data Scientists Make Sense of Model Behavior. *ACM Trans. Comput.-Hum. Interact.* 30, 1 (2023). https://doi.org/10.1145/3542921

[27] Shiye Cao, Anqi Liu, and Chien-Ming Huang. 2024. Designing for Appropriate Reliance: The Roles of AI Uncertainty Presentation, Initial User Decision, and User Demographics in AI-Assisted Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (2024). https://doi.org/10.1145/3637318

[28] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), 1–15. https://doi.org/10.1145/3290605.3300733 Publisher: Association for Computing Machinery.

[29] Ana Caraban, Loukas Konstantinou, and Evangelos Karapanos. 2020. The Nudge Deck: A Design Support Tool for Technology-Mediated Nudging. *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (2020), 395–406. https://doi.org/10.1145/3357236.3395485 Publisher: Association for Computing Machinery.

[30] John M. Carroll and Judith Reitman Olson. 1988. Mental Models in Human-Computer Interaction. In *Handbook of Human-Computer Interaction*, MARTIN HELANDER (Ed.). North-Holland, Amsterdam, 45–65. https://doi.org/10.1016/B978-0-444-70536-5.50007-5

[31] Serena Chen, Kimberly Duckworth, and Shelly Chaiken. 1999. Motivated Heuristic and Systematic Processing. *Psychological Inquiry* 10, 1 (1999), 44–49. http://www.jstor.org/stable/1449522

[32] Xiaogang Chen, Libo Su, and Darrell Carpenter. 2020. Impacts of Situational Factors on Consumers' Adoption of Mobile Payment Services: A Decision-Biases Perspective. *International Journal of Human–Computer Interaction* 36, 11 (2020), 1085–1093. https://doi.org/10.1080/10447318.2020.1722400

[33] Xiaojiao Chen, Xiaoteng Tang, Ying Zhao, Tengyu Huang, Ran Qian, Jiayi Zhang, Wei Chen, and Xiaosong Wang. 2024. Evaluating Visual Consistency of Icon Usage in Across-Devices. *International Journal of Human–Computer Interaction* 40, 9 (2024), 2415–2431. https://doi.org/10.1080/10447318.2022.2162275 Publisher: Taylor & Francis.

[34] Fu-Yin Cherng, Jingchao Fang, Yinhao Jiang, Xin Chen, Taejun Choi, and Hao-Chuan Wang. 2022. Understanding Social Influence in Collective Product Ratings Using Behavioral and Cognitive Metrics. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3517726 Publisher: Association for Computing Machinery.

[35] Chun-Wei Chiang, Zhuoran Lu, Zhuoyan Li, and Ming Yin. 2023. Are Two Heads Better Than One in AI-Assisted Decision Making? Comparing the Behavior and Performance of Groups and Individuals in Human-AI Collaborative Recidivism Risk Assessment. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3581015 Publisher: Association for Computing Machinery.

[36] Isaac Cho, Ryan Wesslen, Alireza Karduni, Sashank Santhanam, Samira Shaikh, and Wenwen Dou. 2017. The Anchoring Effect in Decision-Making with Visual Analytics. *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2017), 116–126. https://doi.org/10.1109/VAST.2017.8585665

[37] Michael Chromik, Malin Eiband, Felicitas Buchner, Adrian Krüger, and Andreas Butz. 2021. I Think I Get Your Point, AI! The Illusion of Explanatory Depth in Explainable AI. *26th International Conference on Intelligent User Interfaces* (2021), 307–317. https://doi.org/10.1145/3397481.3450644 Publisher: Association for Computing Machinery.

[38] Mark Colley, Jan Henry Belz, and Enrico Rukzio. 2021. Investigating the Effects of Feedback Communication of Autonomous Vehicles. *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (2021), 263–273. https://doi.org/10.1145/3409118.3475133 Publisher: Association for Computing Machinery.

[39] Thomas Constant and Guillaume Levieux. 2019. Dynamic Difficulty Adjustment Impact on Players' Confidence. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), 1–12. https://doi.org/10.1145/3290605.3300693 Publisher: Association for Computing Machinery.

[40] John Cook, Stephan Lewandowsky, and Ullrich K. H. Ecker. 2017. Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE* 12, 5 (05 2017), 1–21.

https://doi.org/10.1371/journal.pone.0175799

[41] Florence L. Denmark and Deborah Williams. 2014. *Gender Bias, Overview*. Springer New York, New York, NY, 761–762. https://doi.org/10.1007/978-1-4614-5583-7_430

[42] Evanthia Dimara, Gilles Bailly, Anastasia Bezerianos, and Steven Franconeri. 2019. Mitigating the Attraction Effect with Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 850–860. https://doi.org/10.1109/TVCG.2018.2865233

[43] Evanthia Dimara, Anastasia Bezerianos, and Pierre Dragicevic. 2017. The Attraction Effect in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 471–480. https://doi.org/10.1109/TVCG.2016.2598594

[44] Evanthia Dimara, Steven Franconeri, Catherine Plaisant, Anastasia Bezerianos, and Pierre Dragicevic. 2020. A Task-Based Taxonomy of Cognitive Biases for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics* 26, 2 (2020), 1413–1432. https://doi.org/10.1109/TVCG.2018.2872577

[45] 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. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3517570 Publisher: Association for Computing Machinery.

[46] Tilman Dingler, Benjamin Tag, Evangelos Karapanos, Koichi Kise, and Andreas Dengel. 2020. Workshop on Detection and Design for Cognitive Biases in People and Computing Systems. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3334480.3375159

[47] Tilman Dingler, Benjamin Tag, Philipp Lorenz-Spreen, Andrew W. Vargo, Simon Knight, and Stephan Lewandowsky. 2021. Workshop on Technologies to Support Critical Thinking in an Age of Misinformation. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–5. https://doi.org/10.1145/3411763.3441350

[48] Hongming Dong, Shoufeng Ma, Shuai Ling, Geng Li, Shuxian Xu, and Bo Song. 2023. An Empirical Investigation on the Acceptance of Autonomous Vehicles: Perspective of Drivers' Self–AV Bias. *International Journal of Human–Computer Interaction* 0, 0 (2023), 1–13. https://doi.org/10.1080/10447318.2023.2186000

[49] Tim Donkers and Jürgen Ziegler. 2021. The Dual Echo Chamber: Modeling Social Media Polarization for Interventional Recommending. *Proceedings of the 15th ACM Conference on Recommender Systems* (2021), 12–22. https://doi.org/10.1145/3460231.3474261 Publisher: Association for Computing Machinery.

[50] Tim Draws, Oana Inel, Nava Tintarev, Christian Baden, and Benjamin Timmermans. 2022. Comprehensive Viewpoint Representations for a Deeper Understanding of User Interactions With Debated Topics. *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (2022), 135–145. https://doi.org/10.1145/3498366.3505812 Publisher: Association for Computing Machinery.

[51] Tim Draws, Karthikeyan Natesan Ramamurthy, Ioana Baldini, Amit Dhurandhar, Inkit Padhi, Benjamin Timmermans, and Nava Tintarev. 2023. Explainable Cross-Topic Stance Detection for Search Results. *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (2023), 221–235. https://doi.org/10.1145/3576840.3578296 Publisher: Association for Computing Machinery.

[52] Jessica Maria Echterhoff, Matin Yarmand, and Julian McAuley. 2022. AI-Moderated Decision-Making: Capturing and Balancing Anchoring Bias in Sequential Decision Tasks. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3517443 Publisher: Association for Computing Machinery.

[53] Markus I. Eronen and Laura F. Bringmann. 2021. The Theory Crisis in Psychology: How to Move Forward. *Perspectives on psychological science : a journal of the Association for Psychological Science* 16, 4 (July 2021), 779–788. https://doi.org/10.1177/1745691620970586 Place: United States.

[54] Jonathon St BT Evans and David E Over. 1996. *Rationality and Reasoning*. Psychology Press.

[55] 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.

[56] Jonathan St. B. T. Evans and Keith E. Stanovich. 2013. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science* 8, 3 (2013), 223–241. https://doi.org/10.1177/1745691612460685 arXiv:https://doi.org/10.1177/1745691612460685 PMID: 26172965.

[57] Lauren Fell, Andrew Gibson, Peter Bruza, and Pamela Hoyte. 2020. Human Information Interaction and the Cognitive Predicting Theory of Trust. *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (2020), 145–152. https://doi.org/10.1145/3343413.3377981 Publisher: Association for Computing Machinery.

[58] Cognitive Bias Foundation. 2024. Bias Cheat Sheet. http://bias.transhumanity.net/bias-cheat-sheet/. Accessed: 2024-09-13.

[59] Ujwal Gadiraju, Besnik Fetahu, Ricardo Kawase, Patrick Siehndel, and Stefan Dietze. 2017. Using Worker Self-Assessments for Competence-Based Pre-Selection in Crowdsourcing Microtasks. *ACM Trans. Comput.-Hum. Interact.* 24, 4 (2017). https://doi.org/10.1145/3119930

[60] Christine Geeng, Savanna Yee, and Franziska Roesner. 2020. Fake News on Facebook and Twitter: Investigating How People (Don't) Investigate. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), 1–14. Place: New York, NY, USA Publisher: Association for Computing Machinery.

[61] Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. 2020. A Think-Aloud Study to Understand Factors Affecting Online Health Search. *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (2020), 273–282. https://doi.org/10.1145/3343413.3377961 Publisher: Association for Computing Machinery.

[62] 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

[63] Gerd Gigerenzer. 2008. Why Heuristics Work. *Perspectives on Psychological Science* 3, 1 (2008), 20–29. https://doi.org/10.1111/j.1745-6916.2008.00058.x arXiv:https://doi.org/10.1111/j.1745-6916.2008.00058.x PMID: 26158666.

[64] Thomas Gilovich, Dale Griffin, and Daniel Kahneman. 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press.

[65] Klaus Goffart, Michael Schermann, Christopher Kohl, Jörg Preißinger, and Helmut Krcmar. 2016. Using the Default Option Bias to Influence Decision Making While Driving. *International Journal of Human–Computer Interaction* 32, 1 (2016), 39–50. https://doi.org/10.1080/10447318.2015.1085747

[66] Navita Goyal, Connor Baumler, Tin Nguyen, and Hal Daumé III. 2024. The Impact of Explanations on Fairness in Human-AI Decision-Making: Protected vs Proxy Features. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) *(IUI '24)*. Association for Computing Machinery, New York, NY, USA, 155–180. https://doi.org/10.1145/3640543.3645210

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

[68] Sukeshini A. Grandhi, Linda Plotnick, and Starr Roxanne Hiltz. 2019. Do I Stay or Do I Go? Motivations and Decision Making in Social Media Non-Use and Reversion. *Proc. ACM Hum.-Comput. Interact.* 3, GROUP (2019). https://doi.org/10.1145/3361116

[69] Aditya Gulati, Miguel Angel Lozano, Bruno Lepri, and Nuria Oliver. 2023. BIASeD: Bringing Irrationality into Automated System Design. CEUR. http://hdl.handle.net/10045/132001

[70] Xunhua Guo, Lingli Wang, Mingyue Zhang, and Guoqing Chen. 2023. First Things First? Order Effects in Online Product Recommender Systems. *ACM Trans. Comput.-Hum. Interact.* 30, 1 (2023). https://doi.org/10.1145/3557886

[71] Taehyun Ha and Sangyeon Kim. 2024. 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* (2024), 1–12. https://doi.org/10.1080/10447318.2023.2285640 Publisher: Taylor & Francis.

[72] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. 2022. "Okay, Whatever": An Evaluation of Cookie Consent Interfaces. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3501985 Publisher: Association for Computing Machinery.

[73] Sophia Hadash, Martijn C. Willemsen, Chris Snijders, and Wijnand A. IJsselsteijn. 2022. Improving Understandability of Feature Contributions in Model-Agnostic Explainable AI Tools. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3517650 Publisher: Association for Computing Machinery.

[74] Martin S. Hagger. 2014. Avoiding the "déjà-variable" phenomenon: social psychology needs more guides to constructs. *Frontiers in psychology* 5 (2014), 52. https://doi.org/10.3389/fpsyg.2014.00052 Place: Switzerland.

[75] Ehsan-Ul Haq, Yang K. Lu, and Pan Hui. 2022. It's All Relative! A Method to Counter Human Bias in Crowdsourced Stance Detection of News Articles. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (2022). https://doi.org/10.1145/3555636

[76] MD Romael Haque, Devansh Saxena, Katy Weathington, Joseph Chudzik, and Shion Guha. 2024. Are We Asking the Right Questions?: Designing for Community Stakeholders' Interactions with AI in Policing. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). https://doi.org/10.1145/3613904.3642738 Publisher: Association for Computing Machinery.

[77] Christopher G. Harris. 2019. Detecting Cognitive Bias in a Relevance Assessment Task Using an Eye Tracker. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (2019). https://doi.org/10.1145/3314111.3319824 Publisher: Association for Computing Machinery.

[78] 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

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

[80] Yugo Hayashi, Kosuke Wakabayashi, and Yuki Nishida. 2023. How Sequential Suggestions from a Robot and Human Jury Influence Decision Making: A Large Scale Investigation Using a Court Sentencing Judgment Task. *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (2023), 338–341. https://doi.org/10.1145/3568294.3580101 Publisher: Association for Computing Machinery.

[81] Gaole He, Stefan Buijsman, and Ujwal Gadiraju. 2023. How Stated Accuracy of an AI System and Analogies to Explain Accuracy Affect Human Reliance on the System. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 276 (Oct. 2023), 29 pages. https://doi.org/10.1145/3610067

[82] Gaole He, Lucie Kuiper, and Ujwal Gadiraju. 2023. Knowing About Knowing: An Illusion of Human Competence Can Hinder Appropriate Reliance on AI Systems. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3581025 Publisher: Association for Computing Machinery.

[83] Eric B. Hekler, Predrag Klasnja, Jon E. Froehlich, and Matthew P. Buman. 2013. Mind the Theoretical Gap: Interpreting, Using, and Developing Behavioral Theory in HCI Research. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), 3307–3316. https://doi.org/10.1145/2470654.2466452 Publisher: Association for Computing Machinery.

[84] Sarita Herse, Jonathan Vitale, and Mary-Anne Williams. 2023. Using Agent Features to Influence User Trust, Decision Making and Task Outcome during Human-Agent Collaboration. *International Journal of Human–Computer Interaction* 39, 9 (2023), 1740–1761. https://doi.org/10.1080/10447318.2022.2150691 Publisher: Taylor & Francis.

[85] Ralph Hertwig and Till Grüne-Yanoff. 2017. Nudging and Boosting: Steering or Empowering Good Decisions. *Perspectives on Psychological Science* 12, 6 (2017), 973–986. https://doi.org/10.1177/1745691617702496 arXiv:https://doi.org/10.1177/1745691617702496 PMID: 28792862.

[86] Martin Hilbert. 2012. Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychological bulletin* 138, 2 (2012), 211.

[87] Mohamad Hjeij and Arnis Vilks. 2023. A brief history of heuristics: how did research on heuristics evolve? *Humanities and Social Sciences Communications* 10, 1 (Feb. 2023), 64. https://doi.org/10.1057/s41599-023-01542-z

[88] Christoph Hube, Besnik Fetahu, and Ujwal Gadiraju. 2019. Understanding and Mitigating Worker Biases in the Crowdsourced Collection of Subjective Judgments. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), 1–12. https://doi.org/10.1145/3290605.3300637 Publisher: Association for Computing Machinery.

[89] Jessica Hullman, Eytan Adar, and Priti Shah. 2011. The impact of social information on visual judgments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1461–1470. https://doi.org/10.1145/1978942.1979157

[90] David G. Jansson and Steven M. Smith. 1991. Design fixation. *Design Studies* 12, 1 (1991), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F

[91] Gabbrielle M. Johnson. 2021. Algorithmic bias: on the implicit biases of social technology. *Synthese* 198, 10 (Oct. 2021), 9941–9961. https://doi.org/10.1007/s11229-020-02696-y

[92] 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

[93] 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.

[94] Emma B. Jones and Louise Sharpe. 2017. Cognitive bias modification: A review of meta-analyses. *Journal of Affective Disorders* 223 (2017), 175–183. https://doi.org/10.1016/j.jad.2017.07.034

[95] Daniel Kahneman. 2011. *Thinking, Fast and Slow.* Macmillan.

[96] Daniel Kahneman, Olivier Sibony, and Cass R Sunstein. 2021. *Noise: A flaw in human judgment.* Hachette UK.

[97] Daniel Kahneman and Amos Tversky. 1972. Subjective probability: A judgment of representativeness. *Cognitive psychology* 3, 3 (1972), 430–454.

[98] Daniel Kahneman and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 2 (1979), 263–291. http://www.jstor.org/stable/1914185

[99] Naomi Kakoschke, Rowan Page, Barbora de Courten, Antonio Verdejo-Garcia, and Jon McCormack. 2021. Brain training with the body in mind: Towards gamified approach-avoidance training using virtual reality. *International Journal of Human-Computer Studies* 151 (2021), 102626. https://doi.org/10.1016/j.ijhcs.2021.102626

[100] Alireza Karduni, Isaac Cho, Ryan Wesslen, Sashank Santhanam, Svitlana Volkova, Dustin L Arendt, Samira Shaikh, and Wenwen Dou. 2019. Vulnerable to Misinformation? Verifi! *Proceedings of the 24th International Conference on Intelligent User Interfaces* (2019), 312–323. https://doi.org/10.1145/3301275.3302320 Publisher: Association for Computing Machinery.

[101] Harmanpreet Kaur, Matthew R. Conrad, Davis Rule, Cliff Lampe, and Eric Gilbert. 2024. Interpretability Gone Bad: The Role of Bounded Rationality in How Practitioners Understand Machine Learning. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 77 (apr 2024), 34 pages. https://doi.org/10.1145/3637354

[102] Jieun Kim, Hokyoung Ryu, and Hyeonah Kim. 2013. To be biased or not to be: choosing between design fixation and design intentionality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 349–354. https://doi.org/10.1145/2468356.2468418

[103] Taenyun Kim and Hayeon Song. 2023. Communicating the Limitations of AI: The Effect of Message Framing and Ownership on Trust in Artificial Intelligence. *International Journal of Human–Computer Interaction* 39, 4 (2023), 790–800. https://doi.org/10.1080/10447318.2022.2049134

[104] Tomáš Kliegr, Štěpán Bahník, and Johannes Fürnkranz. 2021. A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. *Artificial Intelligence* 295 (2021), 103458. https://doi.org/10.1016/j.artint.2021.103458

[105] Sara Klüber, Franzisca Maas, David Schraudt, Gina Hermann, Oliver Happel, and Tobias Grundgeiger. 2020. Experience Matters: Design and Evaluation of an Anesthesia Support Tool Guided by User Experience Theory. *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (2020), 1523–1535. https://doi.org/10.1145/3357236.3395552 Publisher: Association for Computing Machinery.

[106] Silvia Knobloch-Westerwick, Benjamin K. Johnson, Nathaniel A. Silver, and Axel Westerwick. 2015. Science Exemplars in the Eye of the Beholder: How Exposure to Online Science Information Affects Attitudes. *Science Communication* 37, 5 (2015), 575–601. https://doi.org/10.1177/1075547015596367 arXiv:https://doi.org/10.1177/1075547015596367

[107] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2018. Frames and Slants in Titles of Visualizations on Controversial Topics. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), 1–12. https://doi.org/10.1145/3173574.3174012 Publisher: Association for Computing Machinery.

[108] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2019. Trust and Recall of Information across Varying Degrees of Title-Visualization Misalignment. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), 1–13. https://doi.org/10.1145/3290605.3300576 Publisher: Association for Computing Machinery.

[109] Loukas Konstantinou, Dionysis Panos, and Evangelos Karapanos. 2024. Exploring the Design of Technology-Mediated Nudges for Online Misinformation. *International Journal of Human–Computer Interaction* (2024), 1–28. https://doi.org/10.1080/10447318.2023.2301265 Publisher: Taylor & Francis.

[110] Thomas Kosch, Robin Welsch, Lewis Chuang, and Albrecht Schmidt. 2023. The Placebo Effect of Artificial Intelligence in Human–Computer Interaction. *ACM Trans. Comput.-Hum. Interact.* 29, 6 (2023). https://doi.org/10.1145/3529225

[111] Morgane Koval and Yvonne Jansen. 2022. Do You See What You Mean? Using Predictive Visualizations to Reduce Optimism in Duration Estimates. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3502010 Publisher: Association for Computing Machinery.

[112] Anastasia Kozyreva, Stephan Lewandowsky, and Ralph Hertwig. 2020. Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools. *Psychological Science in the Public Interest* 21, 3 (2020), 103–156. https://doi.org/10.1177/1529100620946707 arXiv:https://doi.org/10.1177/1529100620946707 PMID: 33325331.

[113] Justin Kruger. 1999. Lake Wobegon be gone! The" below-average effect" and the egocentric nature of comparative ability judgments. *Journal of personality and social psychology* 77, 2 (1999), 221.

[114] Minae Kwon, Erdem Biyik, Aditi Talati, Karan Bhasin, Dylan P. Losey, and Dorsa Sadigh. 2020. When Humans Aren't Optimal: Robots That Collaborate with Risk-Aware Humans. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (2020), 43–52. https://doi.org/10.1145/3319502.3374832 Publisher: Association for Computing Machinery.

[115] Mitra Lashkari and Jinghui Cheng. 2023. "Finding the Magic Sauce": Exploring Perspectives of Recruiters and Job Seekers on Recruitment Bias and Automated Tools. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3581548 Publisher: Association for Computing Machinery.

[116] Nguyen-Thinh Le and Laura Wartschinski. 2018. A Cognitive Assistant for improving human reasoning skills. *International Journal of Human-Computer Studies* 117 (2018), 45–54. https://doi.org/10.1016/j.ijhcs.2018.02.005

[117] Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. 2006. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International Journal of Human-Computer Studies* 64, 10 (2006), 962–973. https://doi.org/10.1016/j.ijhcs.2006.05.002

[118] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2011. Mining Behavioral Economics to Design Persuasive Technology for Healthy Choices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), 325–334. https://doi.org/10.1145/1978942.1978989 Publisher: Association for Computing Machinery.

[119] 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

[120] 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.

[121] Tony W Li, Arshia Arya, and Haojian Jin. 2024. Redesigning Privacy with User Feedback: The Case of Zoom Attendee Attention Tracking. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). https://doi.org/10.1145/3613904.3642594 Publisher: Association for Computing Machinery.

[122] Q. Vera Liao and Wai-Tat Fu. 2014. Can you hear me now? mitigating the echo chamber effect by source position indicators. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 184–196. https://doi.org/10.1145/2531602.2531711

[123] 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. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), 1439–1448. https://doi.org/10.1145/2702123.2702570 Publisher: Association for Computing Machinery.

[124] 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.

[125] Jiqun Liu. 2023. Toward A Two-Sided Fairness Framework in Search and Recommendation. *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (2023), 236–246. https://doi.org/10.1145/3576840.3578332 Publisher: Association for Computing Machinery.

[126] Qianyu Liu, Haoran Jiang, Zihao Pan, Qiushi Han, Zhenhui Peng, and Quan Li. 2024. BiasEye: A Bias-Aware Real-time Interactive Material Screening System for Impartial Candidate Assessment. *Proceedings of the 29th International Conference on Intelligent User Interfaces* (2024), 325–343. https://doi.org/10.1145/3640543.3645166 Publisher: Association for Computing Machinery.

[127] Meagan B. Loerakker, Jasmin Niess, Marit Bentvelzen, and Paweł W. Woźniak. 2024. Designing Data Visualisations for Self-Compassion in Personal Informatics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4 (2024), Article 169.

[128] Shuai Ma, Ying Lei, Xinru Wang, Chengbo Zheng, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2023. Who Should I Trust: AI or Myself? Leveraging Human and AI Correctness Likelihood to Promote Appropriate Trust in AI-Assisted Decision-Making. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3581058 Publisher: Association for Computing Machinery.

[129] Shuai Ma, Xinru Wang, Ying Lei, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2024. "Are You Really Sure?" Understanding the Effects of Human Self-Confidence Calibration in AI-Assisted Decision Making. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). https://doi.org/10.1145/3613904.3642671 Publisher: Association for Computing Machinery.

[130] Peter G. Mahon and Roxanne L. Canosa. 2012. Prisoners and chickens: gaze locations indicate bounded rationality. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 401–404. https://doi.org/10.1145/2168556.2168647

[131] Prateek Mantri, Hariharan Subramonyam, Audrey L. Michal, and Cindy Xiong. 2023. How Do Viewers Synthesize Conflicting Information from Data Visualizations? *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 1005–1015. https://doi.org/10.1109/TVCG.2022.3209467

[132] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (2019). https://doi.org/10.1145/3359183

[133] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021). https://doi.org/10.1145/3411764.3445610 Publisher: Association for Computing Machinery.

[134] Gonzalo Mendez, Luis Galárraga, and Katherine Chiluiza. 2021. Showing Academic Performance Predictions during Term Planning: Effects on Students' Decisions, Behaviors, and Preferences. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021). https://doi.org/10.1145/3411764.3445718 Publisher: Association for Computing Machinery.

[135] Ronald A. Metoyer, Tee Chuanromanee, Gina M. Girgis, Qiyu Zhi, and Eleanor C. Kinyon. 2020. Supporting Storytelling With Evidence in Holistic Review Processes: A Participatory Design Approach. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (2020). https://doi.org/10.1145/3392870

[136] Thomas Mildner, Albert Inkoom, Rainer Malaka, and Jasmin Niess. 2024. Hell is Paved with Good Intentions: The Intricate Relationship Between Cognitive Biases and Dark Patterns. arXiv:2405.07378 [cs.HC] https://arxiv.org/abs/2405.07378

[137] David Moher, Alessandro Liberati, Jennifer Tetzlaff, and Douglas Altman. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *Br Med J* 8 (07 2009), 336–341. https://doi.org/10.1371/journal.pmedl000097

[138] Zachary Munn, Micah D J Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur, and Edoardo Aromataris. 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* 18, 1 (Nov. 2018), 143.

[139] Mohammad Naiseh, Dena Al-Thani, Nan Jiang, and Raian Ali. 2023. How the different explanation classes impact trust calibration: The case of clinical decision support systems. *International Journal of Human-Computer Studies* 169 (2023), 102941. https://doi.org/10.1016/j.ijhcs.2022.102941

[140] Arpit Narechania, Adam Coscia, Emily Wall, and Alex Endert. 2022. Lumos: Increasing Awareness of Analytic Behavior during Visual Data Analysis. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 1009–1018. https://doi.org/10.1109/TVCG.2021.3114827

[141] Feng Ni, David Arnott, and Shijia Gao. 2019. The anchoring effect in business intelligence supported decision-making. *Journal of Decision Systems* 28, 2 (2019), 67–81. https://doi.org/10.1080/12460125.2019.1620573 arXiv:https://doi.org/10.1080/12460125.2019.1620573

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

[143] Mahsan Nourani, Chiradeep Roy, Jeremy E Block, Donald R Honeycutt, Tahrima Rahman, Eric Ragan, and Vibhav Gogate. 2021. Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. *26th International Conference on Intelligent User Interfaces* (2021), 340–350. https://doi.org/10.1145/3397481.3450639 Publisher: Association for Computing Machinery.

[144] Alamir Novin and Eric Meyers. 2017. Making Sense of Conflicting Science Information: Exploring Bias in the Search Engine Result Page. *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (2017), 175–184. https://doi.org/10.1145/3020165.3020185 Publisher: Association for Computing Machinery.

[145] 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. *2016 European Intelligence and Security Informatics Conference (EISIC)* (2016), 148–151. https://doi.org/10.1109/EISIC.2016.038

[146] Tobias Nyström and Moyen M. Mustaquim. 2015. Managing Framing Effects in Persuasive Design for Sustainability. *Proceedings of the 19th International Academic Mindtrek Conference* (2015), 122–129. https://doi.org/10.1145/2818187.2818277 Publisher: Association for Computing Machinery.

[147] 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.

[148] Tadashi Okoshi, Wataru Sasaki, and Jin Nakazawa. 2020. Behavification: bypassing human's attentional and cognitive systems for automated behavior change. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers* (Virtual Event, Mexico) *(UbiComp/ISWC '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 692–695. https://doi.org/10.1145/3410530.3414439

[149] Shaul Oreg and Mahmut Bayazit. 2009. Prone to Bias: Development of a Bias Taxonomy from an Individual Differences Perspective. *Review of General Psychology* 13, 3 (2009), 175–193. https://doi.org/10.1037/a0015656 arXiv:https://doi.org/10.1037/a0015656

[150] Steffi Paepcke and Leila Takayama. 2010. Judging a Bot by Its Cover: An Experiment on Expectation Setting for Personal Robots. *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction* (2010), 45–52. Publisher: IEEE Press.

[151] Marvin Pafla, Kate Larson, and Mark Hancock. 2024. Unraveling the Dilemma of AI Errors: Exploring the Effectiveness of Human and Machine Explanations for Large Language Models. *Proceedings of the CHI Conference on Human Factors in*

Computing Systems (2024). https://doi.org/10.1145/3613904.3642934 Publisher: Association for Computing Machinery.

[152] Charlie Pinder, Jo Vermeulen, Benjamin R. Cowan, and Russell Beale. 2018. Digital Behaviour Change Interventions to Break and Form Habits. *ACM Trans. Comput.-Hum. Interact.* 25, 3 (2018). https://doi.org/10.1145/3196830

[153] Suppanut Pothirattanachaikul, Takehiro Yamamoto, Yusuke Yamamoto, and Masatoshi Yoshikawa. 2020. Analyzing the Effects of "People Also Ask" on Search Behaviors and Beliefs. *Proceedings of the 31st ACM Conference on Hypertext and Social Media* (2020), 101–110. https://doi.org/10.1145/3372923.3404786 Publisher: Association for Computing Machinery.

[154] Marianne Procopio, Ab Mosca, Carlos Scheidegger, Eugene Wu, and Remco Chang. 2022. Impact of Cognitive Biases on Progressive Visualization. *IEEE Transactions on Visualization and Computer Graphics* 28, 9 (2022), 3093–3112. https://doi.org/10.1109/TVCG.2021.3051013

[155] Charvi Rastogi, Yunfeng Zhang, Dennis Wei, Kush R. Varshney, Amit Dhurandhar, and Richard Tomsett. 2022. Deciding Fast and Slow: The Role of Cognitive Biases in AI-Assisted Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1 (2022). https://doi.org/10.1145/3512930

[156] Karen Renaud and Verena Zimmermann. 2018. Ethical guidelines for nudging in information security & privacy. *International Journal of Human-Computer Studies* 120 (2018), 22–35. https://doi.org/10.1016/j.ijhcs.2018.05.011

[157] Eugenia Ha Rim Rho, Gloria Mark, and Melissa Mazmanian. 2018. Fostering Civil Discourse Online: Linguistic Behavior in Comments of #MeToo Articles across Political Perspectives. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (2018). https://doi.org/10.1145/3274416

[158] 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. *Proceedings of the 32nd ACM Conference on Hypertext and Social Media* (2021), 189–199. https://doi.org/10.1145/3465336.3475101 Publisher: Association for Computing Machinery.

[159] Alisa Rieger, Qurat-Ul-Ain Shaheen, Carles Sierra, Mariet Theune, and Nava Tintarev. 2022. Towards Healthy Engagement with Online Debates: An Investigation of Debate Summaries and Personalized Persuasive Suggestions. *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (2022), 192–199. https://doi.org/10.1145/3511047.3537692 Publisher: Association for Computing Machinery.

[160] Lee Ross. 1977. The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process. Advances in Experimental Social Psychology, Vol. 10. Academic Press, 173–220. https://doi.org/10.1016/S0065-2601(08)60357-3

[161] Ulrich Schimmack. 2020. A Meta-Scientific Perspective on "Thinking: Fast and Slow. https://replicationindex.com/2020/12/30/a-meta-scientific-perspective-on-thinking-fast-and-slow/. Accessed: 2024-11-08.

[162] Christina Schwind, Jürgen Buder, and Friedrich W. Hesse. 2011. I Will Do It, but i Don't like It: User Reactions to Preference-Inconsistent Recommendations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), 349–352. https://doi.org/10.1145/1978942.1978992 Publisher: Association for Computing Machinery.

[163] Nikhil Sharma, Q. Vera Liao, and Ziang Xiao. 2024. Generative Echo Chamber? Effect of LLM-Powered Search Systems on Diverse Information Seeking. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). https://doi.org/10.1145/3613904.3642459 Publisher: Association for Computing Machinery.

[164] 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. *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (2023), 107–116. https://doi.org/10.1145/3576840.3578290 Publisher: Association for Computing Machinery.

[165] Wenxuan Wendy Shi, Sneha R. Krishna Kumaran, Hari Sundaram, and Brian P. Bailey. 2023. The Value of Activity Traces in Peer Evaluations: An Experimental Study. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1 (2023). https://doi.org/10.1145/3579627

[166] 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.

[167] Saniat Javid Sohrawardi, Y. Kelly Wu, Andrea Hickerson, and Matthew Wright. 2024. Dungeons & Deepfakes: Using scenario-based role-play to study journalists' behavior towards using AI-based verification tools for video content. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). https://doi.org/10.1145/3613904.3641973 Publisher: Association for Computing Machinery.

[168] Jack B. Soll, Katherine L. Milkman, and John W. Payne. 2015. *A User's Guide to Debiasing.* John Wiley & Sons, Ltd, Chapter 33, 924–951. https://doi.org/10.1002/9781118468333.ch33 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118468333.ch33

[169] Jacob Solomon. 2014. Customization Bias in Decision Support Systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2014), 3065–3074. https://doi.org/10.1145/2556288.2557211 Publisher: Association for Computing Machinery.

[170] Paul M. Spengler, Douglas C. Strohmer, David N. Dixon, and Victoria A. Shivy. 1995. A Scientist-Practitioner Model of Psychological Assessment: Implications for Training, Practice and Research. *The Counseling Psychologist* 23, 3 (1995), 506–534. https://doi.org/10.1177/0011000095233009 arXiv:https://doi.org/10.1177/0011000095233009

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

[172] Keith E. Stanovich and Richard F. West. 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23, 5 (2000), 645–665. https://doi.org/10.1017/S0140525X00003435

[173] Poorna Talkad Sukumar, Ronald Metoyer, and Shuai He. 2018. Making a Pecan Pie: Understanding and Supporting The Holistic Review Process in Admissions. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (2018). https://doi.org/10.1145/3274438

[174] Lu Sun, Hengyuan Zhang, Enze Liu, Mingyang Liu, and Kristen Vaccaro. 2024. NewsGuesser: Using Curiosity to Reduce Selective Exposure. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (2024). https://doi.org/10.1145/3637376

[175] Cass R Sunstein. 2015. Nudging and choice architecture: Ethical considerations. *Yale Journal on Regulation, Forthcoming* (2015).

[176] 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

[177] Pang Suwanaposee, Carl Gutwin, Zhe Chen, and Andy Cockburn. 2023. 'Specially For You' – Examining the Barnum Effect's Influence on the Perceived Quality of System Recommendations. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3580656 Publisher: Association for Computing Machinery.

[178] Siddharth Swaroop, Zana Buçinca, Krzysztof Z. Gajos, and Finale Doshi-Velez. 2024. Accuracy-Time Tradeoffs in AI-Assisted Decision Making under Time Pressure. *Proceedings of the 29th International Conference on Intelligent User Interfaces* (2024), 138–154. https://doi.org/10.1145/3640543.3645206 Publisher: Association for Computing Machinery.

[179] Maxwell Szymanski, Martijn Millecamp, and Katrien Verbert. 2021. Visual, Textual or Hybrid: The Effect of User Expertise on Different Explanations. *26th International Conference on Intelligent User Interfaces* (2021), 109–119. https://doi.org/10.1145/3397481.3450662 Publisher: Association for Computing Machinery.

[180] Richard H. Thaler. 2018. Nudge, not sludge. *Science* 361, 6401 (2018), 431–431. https://doi.org/10.1126/science.aau9241 arXiv:https://www.science.org/doi/pdf/10.1126/science.aau9241

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

[182] Georgios Theocharous, Jennifer Healey, Sridhar Mahadevan, and Michele Saad. 2019. Personalizing with Human Cognitive Biases. *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization* (2019), 13–17. https://doi.org/10.1145/3314183.3323453 Publisher: Association for Computing Machinery.

[183] Ian Thomas, Song Young Oh, and Danielle Albers Szafir. 2024. Assessing User Trust in Active Learning Systems: Insights from Query Policy and Uncertainty Visualization. *Proceedings of the 29th International Conference on Intelligent User Interfaces* (2024), 772–786. https://doi.org/10.1145/3640543.3645207 Publisher: Association for Computing Machinery.

[184] Paul Thomas, Gabriella Kazai, Ryen White, and Nick Craswell. 2022. The Crowd is Made of People: Observations from Large-Scale Crowd Labelling. *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (2022), 25–35. https://doi.org/10.1145/3498366.3505815 Publisher: Association for Computing Machinery.

[185] Suzanne Tolmeijer, Markus Christen, Serhiy Kandul, Markus Kneer, and Abraham Bernstein. 2022. Capable but Amoral? Comparing AI and Human Expert Collaboration in Ethical Decision Making. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3517732 Publisher: Association for Computing Machinery.

[186] Natanael Bandeira Romão Tomé, Madison Klarkowski, Carl Gutwin, Cody Phillips, Regan L. Mandryk, and Andy Cockburn. 2020. Risking Treasure: Testing Loss Aversion in an Adventure Game. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (2020), 306–320. https://doi.org/10.1145/3410404.3414250 Publisher: Association for Computing Machinery.

[187] 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

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

[189] Amos Tversky and Daniel Kahneman. 1988. Rational choice and the framing of decisions. *Decision making: Descriptive, normative, and prescriptive interactions* (1988), 167–192.

[190] Niels van Berkel, Maura Bellio, Mikael B. Skov, and Ann Blandford. 2023. Measurements, Algorithms, and Presentations of Reality: Framing Interactions with AI-Enabled Decision Support. *ACM Trans. Comput.-Hum. Interact.* 30, 2 (2023).

https://doi.org/10.1145/3571815

[191] Elizabeth S. Veinott, James Leonard, Elizabeth Lerner Papautsky, Brandon Perelman, Aleksandra Stankovic, Jared Lorince, Jared Hotaling, Travis Ross, Peter Todd, Edward Castronova, Jerome Busemeyer, Christopher Hale, Richard Catrambone, Elizabeth Whitaker, Olivia Fox, John Flach, and Robert R. Hoffman. 2013. The effect of camera perspective and session duration on training decision making in a serious video game. *2013 IEEE International Games Innovation Conference (IGIC)* (2013), 256–262. https://doi.org/10.1109/IGIC.2013.6659170

[192] Arnav Verma, Luiz Morais, Pierre Dragicevic, and Fanny Chevalier. 2023. Designing Resource Allocation Tools to Promote Fair Allocation: Do Visualization and Information Framing Matter? *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3580739 Publisher: Association for Computing Machinery.

[193] Julien Veytizou, David Bertolo, Charlotte Baraudon, Alexis Olry, and Stéphanie Fleck. 2018. Could a Tangible Interface Help a Child to Weigh His/Her Opinion on Usability? *Proceedings of the 30th Conference on l'Interaction Homme-Machine* (2018), 12–19. https://doi.org/10.1145/3286689.3286702 Publisher: Association for Computing Machinery.

[194] Peter B.M. Vranas. 2000. Gigerenzer's normative critique of Kahneman and Tversky. *Cognition* 76, 3 (2000), 179–193. https://doi.org/10.1016/S0010-0277(99)00084-0

[195] Samangi Wadinambiarachchi, Ryan M. Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 380, 18 pages. https://doi.org/10.1145/3613904.3642919

[196] Emily Wall, Arup Arcalgud, Kuhu Gupta, and Andrew Jo. 2019. A Markov Model of Users' Interactive Behavior in Scatterplots. *2019 IEEE Visualization Conference (VIS)* (2019), 81–85. https://doi.org/10.1109/VISUAL.2019.8933779

[197] Emily Wall, Leslie M. Blaha, Lyndsey Franklin, and Alex Endert. 2017. Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics. *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2017), 104–115. https://doi.org/10.1109/VAST.2017.8585669

[198] Emily Wall, Arpit Narechania, Adam Coscia, Jamal Paden, and Alex Endert. 2022. Left, Right, and Gender: Exploring Interaction Traces to Mitigate Human Biases. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 966–975. https://doi.org/10.1109/TVCG.2021.3114862

[199] Emily Wall, John Stasko, and Alex Endert. 2019. Toward a Design Space for Mitigating Cognitive Bias in Vis. *2019 IEEE Visualization Conference (VIS)* (2019), 111–115. https://doi.org/10.1109/VISUAL.2019.8933611

[200] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), 1–15. https://doi.org/10.1145/3290605.3300831 Publisher: Association for Computing Machinery.

[201] P.C. Wason and J.ST.B.T. Evans. 1974. Dual processes in reasoning? *Cognition* 3, 2 (1974), 141–154. https://doi.org/10.1016/0010-0277(74)90017-1

[202] Martin Wenglinsky. 2017. A Meta-Scientific Perspective on "Thinking: Fast and Slow. https://www.wenglinskyreview.com/wenglinsky-review-a-journal-of-culture-politics/2017/1/23/kahnemans-fallacies. Accessed: 2024-11-08.

[203] Gregory Wheeler. 2020. Bounded Rationality. In *The Stanford Encyclopedia of Philosophy* (Fall 2020 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.

[204] Elizabeth Whitaker, Ethan Trewhitt, Matthew Holtsinger, Christopher Hale, Elizabeth Veinott, Chris Argenta, and Richard Catrambone. 2013. The effectiveness of intelligent tutoring on training in a video game. *2013 IEEE International Games Innovation Conference (IGIC)* (2013), 267–274. https://doi.org/10.1109/IGIC.2013.6659157

[205] Daniel T. Willingham. 2008. Critical Thinking: Why Is It So Hard to Teach? *Arts Education Policy Review* 109, 4 (2008), 21–32. https://doi.org/10.3200/AEPR.109.4.21-32 arXiv:https://doi.org/10.3200/AEPR.109.4.21-32

[206] Timothy D Wilson and Nancy Brekke. 1994. Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations. *Psychological bulletin* 116, 1 (1994), 117.

[207] Timothy D. Wilson, David B. Centerbar, and Nancy Brekke. 2002. *Mental Contamination and the Debiasing Problem.* Cambridge University Press, 185–200.

[208] Luyan Xu, Mengdie Zhuang, and Ujwal Gadiraju. 2021. How Do User Opinions Influence Their Interaction With Web Search Results? *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (2021), 240–244. https://doi.org/10.1145/3450613.3456824 Publisher: Association for Computing Machinery.

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

[210] Mingzhe Yang, Hiromi Arai, Naomi Yamashita, and Yukino Baba. 2024. Fair Machine Guidance to Enhance Fair Decision Making in Biased People. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 285, 18 pages. https://doi.org/10.1145/3613904.3642627

[211] Liudmila Zavolokina, Kilian Sprenkamp, Zoya Katashinskaya, Daniel Gordon Jones, and Gerhard Schwabe. 2024. Think Fast, Think Slow, Think Critical: Designing an Automated Propaganda Detection Tool. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). https://doi.org/10.1145/3613904.3642805 Publisher: Association for Computing Machinery.

[212] Weiyu Zhang, Tian Yang, and Simon Tangi Perrault. 2021. Nudge for Reflection: More Than Just a Channel to Political Knowledge. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021). https://doi.org/10.1145/3411764.3445274 Publisher: Association for Computing Machinery.

[213] Yunfeng Zhang, Rachel K.E. Bellamy, and Wendy A. Kellogg. 2015. Designing Information for Remediating Cognitive Biases in Decision-Making. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), 2211–2220. https://doi.org/10.1145/2702123.2702239 Publisher: Association for Computing Machinery.

[214] Yu Zhang, Jingwei Sun, Li Feng, Cen Yao, Mingming Fan, Liuxin Zhang, Qianying Wang, Xin Geng, and Yong Rui. 2024. See Widely, Think Wisely: Toward Designing a Generative Multi-agent System to Burst Filter Bubbles. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). https://doi.org/10.1145/3613904.3642545 Publisher: Association for Computing Machinery.

[215] Jason Chen Zhao, Wai-Tat Fu, Hanzhe Zhang, Shengdong Zhao, and Henry Duh. 2015. To Risk or Not to Risk? Improving Financial Risk Taking of Older Adults by Online Social Information. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (2015), 95–104. https://doi.org/10.1145/2675133.2685033 Publisher: Association for Computing Machinery.

[216] Chengbo Zheng, Yuheng Wu, Chuhan Shi, Shuai Ma, Jiehui Luo, and Xiaojuan Ma. 2023. Competent but Rigid: Identifying the Gap in Empowering AI to Participate Equally in Group Decision-Making. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3581131 Publisher: Association for Computing Machinery.

[217] 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

[218] Verena Zimmermann and Karen Renaud. 2021. The Nudge Puzzle: Matching Nudge Interventions to Cybersecurity Decisions. *ACM Trans. Comput.-Hum. Interact.* 28, 1 (2021). https://doi.org/10.1145/3429888