Information Processing and Management 61 (2024) 103672

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Cognitive Biases in Fact-Checking and Their Countermeasures: A Review

Michael Soprano [a,*], Kevin Roitero [a], David La Barbera [a], Davide Ceolin [b], Damiano Spina [c], Gianluca Demartini [d], Stefano Mizzaro [a]

[a] *University of Udine, Via Delle Scienze 206, Udine, Italy*
[b] *Centrum Wiskunde & Informatica (CWI), Science Park 123, Amsterdam, The Netherlands*
[c] *RMIT University, 124 La Trobe St, Melbourne, Australia*
[d] *The University of Queensland, St Lucia QLD 4072, Brisbane, Australia*

ARTICLE INFO

Keywords:
Cognitive bias
Misinformation
Fact-checking
Truthfulness assessment

ABSTRACT

The increase of the amount of misinformation spread every day online is a huge threat to the society. Organizations and researchers are working to contrast this misinformation plague. In this setting, human assessors are indispensable to correctly identify, assess and/or revise the truthfulness of information items, i.e., to perform the fact-checking activity. Assessors, as humans, are subject to systematic errors that might interfere with their fact-checking activity. Among such errors, cognitive biases are those due to the limits of human cognition. Although biases help to minimize the cost of making mistakes, they skew assessments away from an objective perception of information. Cognitive biases, hence, are particularly frequent and critical, and can cause errors that have a huge potential impact as they propagate not only in the community, but also in the datasets used to train automatic and semi-automatic machine learning models to fight misinformation. In this work, we present a review of the cognitive biases which might occur during the fact-checking process. In more detail, inspired by PRISMA – a methodology used for systematic literature reviews – we manually derive a list of 221 cognitive biases that may affect human assessors. Then, we select the 39 biases that might manifest during the fact-checking process, we group them into categories, and we provide a description. Finally, we present a list of 11 countermeasures that can be adopted by researchers, practitioners, and organizations to limit the effect of the identified cognitive biases on the fact-checking activity.

1. Introduction

The amount of information which is generated daily by users on social media platforms, news agencies, and the web in general is rapidly increasing. As a result, organizations performing fact-checking of information items are overwhelmed by the amount

* Corresponding author.

🐦 Michael Soprano, Linkedin: michaelsoprano (M. Soprano), 🐦 David La Barbera, Linkedin: david-la-barbera-8a1a646a (D. La Barbera), 🐦 Davide Ceolin, Linkedin: davideceolin (D. Ceolin), 🐦 Damiano Spina, Linkedin: damianospina (D. Spina), 🐦 Gianluca Demartini, Linkedin: gianlucademartini (G. Demartini), 🐦 Stefano Mizzaro, Linkedin: stefano-mizzaro-1234082 (S. Mizzaro).

E-mail addresses: michael.soprano@uniud.it (M. Soprano), kevin.roitero@uniud.it (K. Roitero), david.labarbera@uniud.it (D. La Barbera), davide.ceolin@cwi.nl (D. Ceolin), damiano.spina@rmit.edu.au (D. Spina), demartini@acm.org (G. Demartini), stefano.mizzaro@uniud.it (S. Mizzaro).

URLs: http://www.michaelsoprano.com (M. Soprano), http://www.kevinroitero.com (K. Roitero), http://www.cwi.nl/en/people/davide-ceolin/ (D. Ceolin), http://www.damianospina.com (D. Spina), http://www.gianlucademartini.net (G. Demartini), http://users.dimi.uniud.it/~stefano.mizzaro/ (S. Mizzaro).

https://doi.org/10.1016/j.ipm.2024.103672
Received 10 July 2023; Received in revised form 13 December 2023; Accepted 22 January 2024
Available online 3 February 2024

of material that requires inspection (Porter, 2020). Moreover, the current processes and pipelines implemented by fact-checking organizations cannot cope with the current trend as they are not designed to scale up to the massive amount of information that requires attention. The problem is so serious that the World Health Organization (WHO) director-general used the neologism "infodemic" to refer to the problem of misinformation during the 2020 edition of the Munich Security Conference,[1] while the COVID-19 pandemic was still ongoing (Alam et al., 2021).

People such as fact-checkers, being experts (Drobnic Holan, 2018; FactCheck.org, 2020; The RMIT ABC Fact Check Team, 2021) or crowd workers (Demartini, Mizzaro, & Spina, 2020; Roitero, Soprano, Fan, et al., 2020; Roitero, Soprano, Portelli, et al., 2020; Soprano et al., 2021), can be subject to systematic errors that can harm the information assessment process. Systematic errors due to the limits of human cognition are called cognitive biases. According to the Oxford Dictionary, a general definition of bias is "a strong feeling in favor of or against one group of people, or one side in an argument, often not based on fair judgment". There exist different types of biases: cognitive biases, conflicts of interest, statistical biases, prejudices. In this paper, we focus on cognitive biases because these are systematic biases due to limits in human cognition that can unintentionally affect the effectiveness of fact-checking processes. From a psychological point of view, evolutionary studies suggest that humans developed biases to minimize the cost of making mistakes in the long period, as they can improve decision-making processes (Johnson, Blumstein, Fowler, & Haselton, 2013). According to error management studies that aim at explaining human processes in decision-making, cognitive biases, defined as "the ones that skew our assessments away from an objective perception of information" (Johnson et al., 2013), have been favored by nature in order to minimize whichever error that caused a great cost (Haselton & Nettle, 2006; Nesse, 2001, 2005). In fact, decision-making processes are often complex, and we are not always capable of keeping up to date – and statistically correct – the estimations of the error probabilities involved in such processes; thus, natural selection might have favored cognitive biases to simplify the overall decision process (Cosmides & Tooby, 1994; Todd & Gigerenzer, 2000). Summarizing, cognitive biases evolved because of the intrinsic limitations of humans when making a decision. Cognitive biases play a major role in the way (mis)information and verified content are consumed (Draws et al., 2022), and different debiasing strategies have been proposed in relation to cognitive factors such as people's memory for misinformation (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012); also, debiasing is not the only possible choice, as other proposals aim at managing bias instead of removing it (Demartini, Roitero, & Mizzaro, 2021).

It is important to remark that biases can have far-fetched consequences. Keeping the focus on fact-checking, machine learning is an interesting potential solution to address the obvious scalability issues of the approach based on human experts (Ciampaglia et al., 2015; Liu & Wu, 2020; Wang, 2017; Weiss & Taskar, 2010). In this respect, biases not only interfere with the human fact-checking activity in practice, but they also create issues for automatic approaches as they creep into the datasets that are then used to train the machine learning systems, in some cases contributing to blatant errors (see, e.g., the famous "Gorilla Case" (Simonite, 2018) where an image recognition algorithm misclassified black people as "gorillas"). Since biases introduce errors due to systematic limits in human cognition that are potentially shared among several individuals, bias prevention, management, and/or control are fundamental.

The contributions of this work are four-fold. First, we propose a comprehensive list of 221 cognitive biases by reviewing the available related literature. Next, we extract the subset of 39 biases that may manifest during the fact-checking activity. Furthermore, we categorize them and propose a list of countermeasures to limit their impact. Lastly, we outline the building blocks of a bias-aware assessment pipeline for fact-checking, with each countermeasure mapped to a constituting block of the pipeline.

The remainder of the paper is structured as follows. Section 2 provides some background on the fact-checking process and on cognitive biases. In Section 3 we discuss aims and motivations of our work, as well as the relations and differences with other existing reviews. Next, we describe the methodology used in this paper (Section 4). We then present the list of cognitive biases selected in this work as relevant to fact-checking (Section 5), along with their categorization (Section 6). Finally, we propose a list of countermeasures (Section 7) and the constituting blocks of a bias-aware assessment pipeline (Section 8) to be employed in fact-checking. To conclude, we discuss the limitations of our approach (Section 9) and provide implications of the work, future work, and a summary of the contributions (Section 10).

2. Background

To contextualize our research within the existing body of work, we describe the fact-checking process by presenting how some notable organizations work in practice (Section 2.1). Next, we investigate in Section 2.2 the methodologies developed by researchers to evaluate the truthfulness of information items (i.e., the core process of the fact-checking activity) using crowdsourcing and machine learning. Then, in Section 2.3, we recall works on cognitive biases and we briefly discuss existing literature reviews and comparing them with the scope of our paper.

2.1. The process of fact-checking

Fact-checking is a complex process that involves several activities (Mena, 2019; Vlachos & Riedel, 2014). An abstract and general pipeline for fact-checking might include the following steps (not necessarily in this order): check-worthiness (i.e., ensure that an

[1] https://www.who.int/director-general/speeches/detail/munich-security-conference

information item is of great interest for a possibly large audience), evidence retrieval (i.e., retrieve the evidence needed to fact-check the information item), truthfulness assessment, discussion among the assessors to reach a consensus, and assignment and publication of the final truthfulness score for the information item inspected. With respect to this pipeline, in this paper we focus mainly on the truthfulness assessment step.

It is also interesting to briefly examine the fact-checking processes adopted in practice by three famous organizations, namely PolitiFact, RMIT ABC Fact Check, and FactCheck.org – verified signatories to the International Fact-Checking Network (IFCN, https://www.poynter.org/ifcn/) – given that they set a de-facto standard for the pipeline required to perform fact-checking at scale.

PolitiFact[2] fact-checks information items by US Politicians. Drobnic Holan (2018) details the process as follows. The reporter in charge of running the fact-checking proposes, to perform the truthfulness assessment step, a rating using a six-level ordinal scale (Pants On Fire, False, Mostly False, Half True, Mostly True, and True). Such assessment is reported to an editor. The reporter and the editor work together to reach a consensus on the rating proposed by adding clarifications and details if needed. Then, the information item is shown to two additional editors, which review the work of the editor and the reporter by providing answer to the following four questions (Drobnic Holan, 2018, Section "How We Determine Truth-O-Meter Ratings"): (1) Is the information item literally true? (2) Is there another way to read the information item? Is the information item open to interpretation? (3) Did the speaker provide evidence? Did the speaker prove the information item to be true? (4) How have we handled similar information items in the past? What is PolitiFact's jurisprudence? Then, the definitive rating of the item is decided upon using the majority vote of the score submitted by the editors, final edits are made to make sure everything is consistent, and the report is finally published.

RMIT ABC Fact Check[3] focuses on information items made by Australian public figures, advocacy groups, and institutions. Their process works as follows (see The RMIT ABC Fact Check Team, 2021 for a more detailed description). The information item to be checked needs to be approved by the director which assesses its check-worthiness. Then, one of the researchers at RMIT ABC Fact Check contacts experts in the field and occasionally the claimant to retrieve evidence and get back data which can be helpful for fact-checking. The researcher writes the data and the information. An expert fact-checker inspects and reviews them. In this stage, the expert fact-checker identifies possible problems and questions the researcher on anything that they might have missed (e.g., missing or not exhaustive evidence retrieved). The expert fact-checker and the researcher revise the draft until the fact-checker is satisfied with the outcome; then, the whole team discusses the final verdict for the item. The final verdict is then expressed on a fine-grained categorical scale, which is used in their publications. For documentation purposes, the verdict is further refined into a three-level ordinal scale defining its truthfulness value: False, In Between, True. This choice is based on previous work demonstrating that a three-level scale may be the most suitable.

FactCheck.org[4] fact-checks information items dealing with US politicians. Their process works as follows (see FactCheck.org, 2020 for a more detailed description). As for the check-worthiness step, they select items made by the president of the United States and important politicians, focusing on those made by presidential candidates during public appearances, top senate races, and congress actions. To perform evidence retrieval, they seek through video transcripts or articles to identify possible misleading or false information items and ask the organization or the person making the item to prove its truthfulness by providing supporting documentation. If no evidence is provided, FactCheck.org searches trusted sources for evidence confirming or refusing the item. Finally, the verdict about the information item is published, without assigning a fine-grained truthfulness label. At FactCheck.org, each item is revised in most cases by four people (FactCheck.org, 2020, Section "Editing"): a line editor (reviewing content), a copy editor (reviewing style and grammar), a fact-checker (in charge of the fact-checking process), and the director of the Annenberg Public Policy Center.

In summary, the fact-checking processes of the three organizations share similarities and differences, described in Table 1. The table reports, for each fact-checking organization considered, the information items provenance, together with the claimants considered, the truthfulness scale used, and the number of expert fact-checkers involved in the process. All three organizations are committed to upholding the principles of the International Fact-Checking Network (IFCN) and focus on checking information items made by politicians and/or public figures. However, they differ in the specific process followed for evidence retrieval, truthfulness assessment, and rating of the items. PolitiFact focuses on US politicians, using a six-level rating scale and a consensus-based process among editors and reporters to determine the final rating. RMIT ABC Fact Check targets Australian public figures, engaging field experts and a collaborative review process where fact-checking is performed by three experts, thus having the whole team decide the final verdict. FactCheck.org also concentrates on US politicians, seeking evidence from claimants and trusted sources, and has a four-person team to review each information item. Despite these differences, all three organizations demonstrate a strong commitment to accuracy, transparency, and thoroughness in their fact-checking processes, providing valuable resources for the public to access reliable information on political items. Moreover, it is worth mentioning that since all three organizations rely exclusively on human judgment for their evaluations, their processes are potentially susceptible to the set of cognitive biases addressed in this paper.

2.2. Fact-checking: Crowdsourcing and machine learning

The process of fact-checking, as implemented by the organizations, has the clear limitation of being not scalable: it requires experts and, being rather time-consuming, is clearly not capable of coping with the huge amount of information shared online

[2] https://politifact.com/

[3] https://www.abc.net.au/news/factcheck

[4] https://www.factcheck.org/

Table 1
Statistics of the fact-checking process as performed by different organizations.

Organization	Country	Claimants	Truthfulness scale	Experts involved
PolitiFact	USA	Politicians	Pants On Fire, False, Mostly False, Half True, Mostly True, True	4
RMIT ABC Fact Check	Australia	Public Figures, Advocacy Groups, Institutions	False, In Between, True	3
FactCheck.org	USA	Politicians	/	4

every day (Das, Liu, Kovatchev, & Lease, 2023; Demartini et al., 2020; Spina et al., 2023). For this reason, the misinformation problem gained attention among researchers who tried to develop alternative methodologies which can help in identifying and evaluating the truthfulness of online information. To this end, essentially two main approaches have been proposed.

Fact-checking has been studied with the help of crowdsourcing techniques, which have been leveraged to obtain reliable labels at scale (Demartini, Difallah, Gadiraju, & Catasta, 2017; Demartini et al., 2020). Draws, Rieger, Inel, Gadiraju, and Tintarev (2021) focused on debiasing strategies for crowdsourcing, proposing a methodology that researchers and practitioners can follow to improve their task design and report limitations of collected data; Eickhoff (2018) gathered labels to study and analyze assessor bias, finding that they have major effects on annotation quality (Eickhoff, 2018); Li et al. (2020) worked in the setting of Twitter topic models; La Barbera, Roitero, Demartini, Mizzaro, and Spina (2020) and Roitero, Demartini, Mizzaro, and Spina (2018), Roitero, Soprano, Fan, et al. (2020) collected thousands of truthfulness labels focusing on crowd workers' effectiveness, agreement, and bias. They show that crowdsourcing labels correlate with expert judgments, and workers' backgrounds and biases (e.g., political orientation) have a major impact on label quality. Roitero, Soprano, Portelli, et al. (2020) worked in a similar setting with a focus on COVID-19 pandemic-related information items. Some other works focused on the credibility and trust of sources of information (Bhuiyan, Zhang, Sehat, & Mitra, 2020; Epstein, Pennycook, & Rand, 2020) and on echo chambers and filter bubbles (Eady, Nagler, Guess, Zilinsky, & Tucker, 2019; Nguyen, 2020).

Besides human-powered systems, other lines of research investigated the usage of automatic machine learning techniques for fact-checking (Hassan et al., 2015; Thorne & Vlachos, 2018). Such techniques rely on training some machine learning algorithm on a labeled dataset which is usually built using human assessors. Thus, the set of biases that are present in the dataset might impact the trained machine learning model (Caliskan, Bryson, & Narayanan, 2017). For this reason, we briefly report related work in the setting of machine learning techniques to cope with disinformation. A set of works focused on the bias of the datasets used to train machine learning models: Vlachos and Riedel (2014) defined the setting and the challenges needed to create a benchmark dataset for fact-checking; Ferreira and Vlachos (2016) described a dataset for stance classification; Wang (2017) created the LIAR dataset which contains a large collection of fact-checked information items; Liu and Wu (2020) used a deep neural network architecture to perform detection and stop the spreading of misinformation before it becomes viral. Another line of research focused not on the data but rather on the algorithms which can be employed to build a fully automatic methodology to fact-check information: Weiss and Taskar (2010) developed a method based on adversarial networks; Ciampaglia et al. (2015) used an approach based on knowledge networks; Alhindi, Petridis, and Muresan (2018) leveraged justification modeling; Reis, Correia, Murai, Veloso, and Benevenuto (2019) and Wu, Rao, Yang, Wang, and Nazir (2020) discussed explainable machine learning algorithms that can be employed for fake news detection; Evans, Edge, Larson, and White (2020) and Oeldorf-Hirsch and DeVoss (2020) considered information sources and their metadata.

2.3. Cognitive biases

Given that the truthfulness assessment step of the fact-checking pipeline is driven by human assessments, either as direct human assessments or when using the human labeled data to train machine learning models, it is prone to suffer from cognitive biases. According to the literature, more than 180 cognitive biases exist (Caverni, Fabre, & Gonzalez, 1990; Haselton, Nettle, & Murray, 2015; Hilbert, 2012; Kahneman & Frederick, 2002). Even if a standard conceptualization or classification of such biases is a debated problem (Gigerenzer & Selten, 2008; Hilbert, 2012), many works confirmed the presence of bias in many domains using reproducible studies (Thomas, 2018), for example in information seeking and retrieval (Azzopardi, 2021). Furthermore, biases are often classified by their generative mechanism (Hilbert, 2012); Oeberst and Imhoff (2023), for instance, argue that multiple biases can be generated by a given fundamental belief. Research also agreed that multiple biases can occur at the same time (MacCoun, 1998; Nickerson, 1998).

It is common knowledge that cognitive biases affect the reasoning process, decision-making, and human behavior in general. Their effect has been widely studied in multiple disciplines: Barnes (1984), Das and Teng (1999), Ehrlinger, Readinger, and Kim (2016), Hilbert (2012), and Swets, Dawes, and Monahan (2000) studied the effect of cognitive biases in decision processes and planning, Fisher and Statman (2000) focused on market forecasting, Draws et al. (2021) and Eickhoff (2018) analyzed cognitive biases in crowdsourcing, Baeza-Yates presented an overview of biases in the web (Baeza-Yates, 2018) and in search and recommendation systems (Baeza-Yates, 2020), Sylvia Chou, Gaysynsky, and Cappella (2020) studied the role of cognitive biases in social media platforms, Kiesel, Spina, Wachsmuth, and Stein (2021) studied biases related to presentation format using conversational interfaces in the context of systems for argument search. All those works reported that bias is strongly correlated with data quality and with the effectiveness of the systems and models detailed in the papers.

Among all the studies dealing with cognitive biases, a line of research has focused on the role of specific cognitive biases in relation to the misinformation topic. Park, Park, and Kang (2021) discuss the fact-checking activity, asserting that its effectiveness varies due to multiple factors. They illustrate how statements that are neither entirely false nor true, often resulting in borderline judgments, can manifest unexpected cognitive biases in human perception. Meanwhile, Zhou and Zafarani (2020) survey and evaluate methods used to detect misinformation from four perspectives. They argue that people's trust in fake news can be built when the fake news confirms one's preexisting political biases (i.e., particular cognitive biases), thus providing resources to evaluate the partisanship of news publishers. Mastroianni and Gilbert (2023) show, in a recent study, how biased exposure to information and biased memory for information makes people believe that morality is declining for decades. Zollo (2019) studied how information spreads across communities on Facebook, focusing on echo chambers and confirmation bias. They provide empirical evidence of echo chambers and filter bubbles, showing that confirmation bias plays a crucial role in content selection and diffusion (Cinelli, Cresci, Quattrociocchi, Tesconi, & Zola, 2022; Cinelli, De Francisci, Galeazzi, Quattrociocchi, & Starnini, 2021; Cinelli et al., 2020; Del Vicario et al., 2016; Zollo & Quattrociocchi, 2018). Wesslen et al. (2019) explored the role of visual anchors in the decision-making process related to Twitter misinformation. They found that these visual anchors significantly impact users in terms of activity, speed, confidence, and accuracy. Karduni et al. (2018) focused on uncertainty on truthfulness assessment when employing visual analysis. Acerbi (2019) analyzed a cognitive attraction phenomenon in online misinformation, identifying a set of cognitive features that contribute to the spread of misinformation. Chou, Gaysynsky, and Vanderpool (2021) investigated factors such as biases driving misinformation sharing and acceptance in the context of COVID-19. Traberg and Van Der Linden (2022) investigated the role of perceived source credibility in mitigating the effects of political bias. Zhou and Shen (2022) considered confirmation bias on misinformation related to the topic of climate change. Ceci and Williams (2020) propose "adversarial fact-checking", i.e., pairing fact-checkers from different sociopolitical backgrounds, as a mechanism to address biases that may occur when verifying political claims.

While the previously cited works focus on specific aspects, there are literature reviews that relate with the issues of cognitive biases and the overall misinformation spreading problem. Ruffo, Semeraro, Giachanou, and Rosso (2023) address the most important psychological effects that provide provisional explanations for reported empirical observations regarding the mechanisms behind the spread of misinformation on social networks. Wang, McKee, Torbica, and Stuckler (2019) reviews works specifically focused on the spread of health-related misinformation, but they explicitly choose to avoid the extensive literature related to cognitive biases. Tucker et al. (2018) reviews research findings related to cognitive biases in political discourse, with a focus on the detection of computational propaganda.

More generally, the literature includes recent reviews that address cognitive biases in various fields of study not related to misinformation and fact-checking. Armstrong et al. (2023) specifically explores biases that may impact surgical events and discusses mitigation strategies used to reduce their effects. Similarly, Vyas, Murphy, and Greenberg (2023) investigates biases affecting military personnel. Eberhard (2023) reviews strategies to mitigate the effects of cognitive biases resulting from visualization strategies on judgment and decision-making. Additionally, Gomroki, Behzadi, Fattahi, and Fadardi (2023) collect data on cognitive biases in information retrieval.

3. Aims & motivations

In addressing the challenges posed by cognitive biases that may impact the fact-checking activity, various studies have focused on different aspects of fact-checking. There is still a gap in comprehensively reviewing how cognitive biases specifically influence the fact-checking process.

Existing literature reviews have primarily concentrated on technical and procedural aspects of fact-checking (Zhou & Zafarani, 2020), addressing cognitive biases partially. For instance, they propose generative mechanisms for subsets of cognitive biases (Oeberst & Imhoff, 2023), address partisanship-related biases exclusively (Walter, Cohen, Lance Holbert, & Morag, 2020), review only those deemed as the most important by the authors (Ruffo et al., 2023), or avoid the aspect completely (Wang et al., 2019). Some reviews focus on the context of political conversations only (Tucker et al., 2018). Other recent reviews dealing with cognitive biases are contributions to other fields of study (Armstrong et al., 2023; Eberhard, 2023; Gomroki et al., 2023; Vyas et al., 2023). The existing reviews, thus, often overlook the underlying set of cognitive biases that can significantly impact the outcomes of the fact-checking activity. Our review is aimed at filling these gaps and we believe it is necessary because cognitive biases are inherent in human judgments and, as a consequence, in the datasets used for training machine learning models for the fact-checking domain. These biases can subtly but profoundly influence the effectiveness and reliability of fact-checking processes. By identifying and understanding these biases, we can develop more robust and unbiased fact-checking methodologies, being them human-based or automatic. This approach not only enhances the accuracy of fact-checking but also contributes to the broader discussion on the reliability and trustworthiness of information.

In this review, our primary motivation is to provide a comprehensive and systematic investigation of the cognitive biases that may manifest during the fact-checking process, compromising its effectiveness in a real-world scenario. Thus, the purpose of this review is fourfold: (i) to systematically identify the cognitive biases that are relevant to the fact-checking process, (ii) to provide a categorization of these biases and real-world examples to illustrate their impact on fact-checking, (iii) to propose potential countermeasures that can help mitigate the risk of cognitive biases manifesting in a fact-checking context, and (iv) to provide the constituting blocks of a bias-aware fact-checking pipeline that helps to minimize such a risk. In more detail, we adopt PRISMA, a methodology well-grounded in the literature, to systematically collect and report biases. Specifically, our focus is on retrieving a single formulation for each considered bias and, if possible, a single reference to support its framing in a fact-checking-related scenario. Our goal, therefore, is not to build a comprehensive list of references, but rather, a comprehensive list of biases.

In summary, our work aims to characterize cognitive biases that may manifest during the fact-checking activity, offering a novel perspective that complements and extends existing literature. While we acknowledge that our proposal is not conclusive and should be regarded as a starting point for further investigation in this area, our aim is to provide valuable insights and practical guidance for researchers, practitioners, and policymakers working in the field of fact-checking and information assessment. Furthermore, to our knowledge, this is the first work to provide a comprehensive set of countermeasures to prevent cognitive biases in the fact-checking process.

4. Methodology

We initially introduce the PRISMA methodology, an approach to conduct high-quality systematic reviews and meta-analyses. Then, we describe how we adopt it to find existing literature about cognitive biases. To summarize, we explore various information sources to find literature that addresses one or more examples of cognitive biases by relying on our search strategy. From the literature found, we perform data collection by extracting one or more biases according to our eligibility criteria. Then, we filter the whole list of cognitive biases according to our selection process to obtain only those cognitive biases that might manifest while performing the fact-checking activity.

4.1. The PRISMA methodology

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. Moher, Liberati, Tetzlaff, and Altman (2009) originally proposed the approach in 2009 and it was a reformulation of the QUORUM guidelines (Moher et al., 1999). In 2020, Page, McKenzie, et al. (2021) proposed an updated version known as "The PRISMA 2020 Statement". In the following, we describe and refer to such a formulation.

PRISMA is a transparent approach that has been widely adopted in various research fields. It aims to help researchers in conducting high-quality systematic reviews and meta-analyses. Its clear and structured framework facilitates the identification, assessment, and synthesis of relevant data, ensuring that the review process is rigorous, replicable, and unbiased. At its core, PRISMA consists of a checklist[5] and a flow diagram,[6] both publicly available.

The PRISMA checklist is composed of 27 items addressing the introduction, methods, results, and discussion sections of a systematic review report, summarized in Table 2. Items 10, 13, 16, 20, 23, and 24 are further split into sub-items (not shown in the table). The flow diagram, on the other hand, depicts the flow of information through the different phases of a systematic review. It maps out the number of records identified, included, and excluded, together with the rationale for exclusions; it is available in two forms, depending on whether the review is a new contribution or an updated version of an existing one. Given that the review we are proposing is a new contribution, we rely on the former version, reported in Fig. 1. If we compare the figure with Table 2, we can see how the diagram provides further details mostly related to items 5 to 10, and 16 of the checklist.

The checklist and the diagram should be used according to the "Explanation and Elaboration Document" (Page, Moher, et al., 2021), which aims to enhance the use, understanding, and dissemination of a review made using PRISMA. Several extensions[7] have been developed to facilitate the reporting of different types or aspects of systematic reviews. The "PRISMA 2020 Extension For Abstracts", published together with the overall statement (Page, McKenzie, et al., 2021), is a 12-item checklist that gives researchers a framework for condensing their systematic reviews into abstracts for journals or conferences. To summarize, the review proposed in this paper is conducted by relying on four main PRISMA elements: (i) the checklist, (ii) the flow diagram, (iii) the checklist for abstracts, and (iv) the explanation and elaboration document of the PRISMA 2020 statement (Page, McKenzie, et al., 2021; Page, Moher, et al., 2021).

Given that our goal is to identify and extract the cognitive biases that might manifest in the fact-checking process from the whole set of cognitive biases described in the literature, rather than finding all the research papers that address cognitive biases to some extent (see Section 3), not all the items provided by the PRISMA checklist have to be addressed. However, we recognize the importance of adhering to the original guidelines as much as possible. Thus, we start by detailing the checklist items that we did not address because they are not relevant for the goal of finding biases or simply because they are not needed. In more detail, we do not: compute effect measures (Item 12), study heterogeneity and robustness of the synthesized results (Item 13), present assessments of risk of bias for each included study (Item 18), perform statistical analyses (Item 20), present assessments of risk of bias due to missing results for each synthesis (Item 21), present assessments of certainty in the body of evidence for each outcome (Item 22), provide particular registration information about the review (Item 24).

Most of the work performed to adopt PRISMA concerns the alterations of the inclusion and exclusion criteria that would typically be applied to research articles in a literature review and the collection and selection process of the cognitive biases presented in

[5] http://prisma-statement.org/PRISMAStatement/Checklist

[6] http://prisma-statement.org/PRISMAStatement/FlowDiagram

[7] http://prisma-statement.org/Extensions/

Table 2
The 27 items of the PRISMA checklist.
Source: Adapted from Page, McKenzie, et al. (2021).

Item #	Section/Topic
1	Title
2	Abstract
3	Rationale
4	Objectives
5	Eligibility Criteria
6	Information Sources
7	Search Strategy
8	Selection Process
9	Data Collection Process
10	Data Items
11	Study Risk Of Bias Assessment
12	Effect Measures
13	Synthesis Methods
14	Reporting Bias Assessment
15	Certainty Assessment
16	Study Selections
17	Study Characteristics
18	Risk Of Bias In Studies
19	Results Of Individual Studies
20	Results Of Syntheses
21	Reporting Biases
22	Certainty Of Evidences
23	Discussion
24	Registration And Protocol
25	Support
26	Competing Interests
27	Availability Of Data

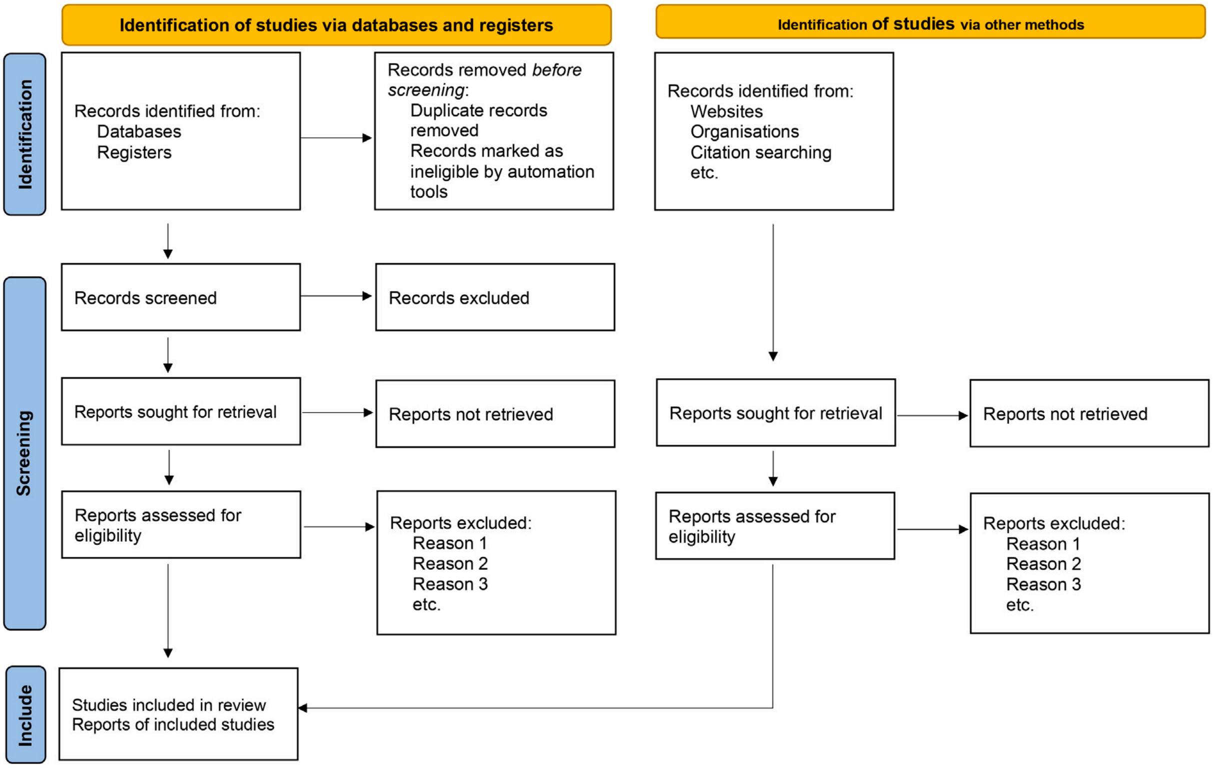

Identification of studies via databases and registers
Identification of studies via other methods
Identification
Screening
Include
Records identified from: Databases Registers
Records removed *before screening*: Duplicate records removed Records marked as ineligible by automation tools
Records identified from: Websites Organisations Citation searching etc.
Records screened
Records excluded
Reports sought for retrieval
Reports not retrieved
Reports sought for retrieval
Reports not retrieved
Reports assessed for eligibility
Reports excluded: Reason 1 Reason 2 Reason 3 etc.
Reports assessed for eligibility
Reports excluded: Reason 1 Reason 2 Reason 3 etc.
Studies included in review Reports of included studies

Fig. 1. The PRISMA flow diagram for new systematic reviews which included searches of databases, registers and other sources.
Source: Adapted from Page, McKenzie, et al. (2021)

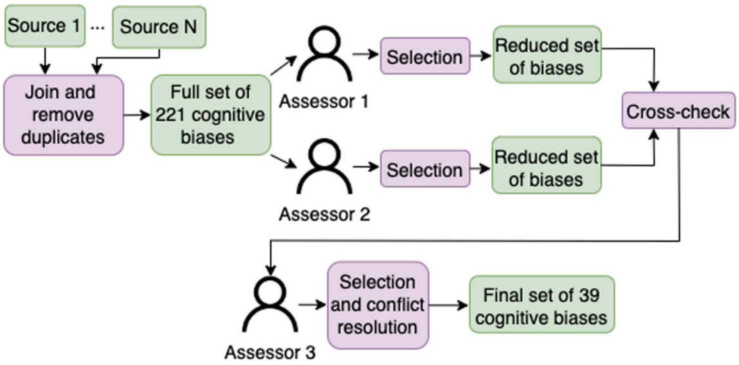

Source 1
...
Source N
Join and remove duplicates
Full set of 221 cognitive biases
Assessor 1
Selection
Reduced set of biases
Assessor 2
Selection
Reduced set of biases
Cross-check
Assessor 3
Selection and conflict resolution
Final set of 39 cognitive biases

Fig. 2. Data collection and selection process.

the following; that is, how we approach the items 5–10 and 13 reported in Table 2 and how we perform the processes described by the flow diagram shown in Fig. 1. This tailored adaptation allows us to follow PRISMA's structured approach – which involves predefined eligibility criteria, search strategies, and data extraction – that helps minimize the risk of errors in our review process by considering a slightly different final outcome. Section 4.2 describes our eligibility criteria, information sources and search strategy, while Section 5 details the data collection and selection processes. The full "PRISMA Abstract Checklist" and "PRISMA Checklist" are reported in Appendix A.

4.2. Eligibility criteria, information sources, and search strategy

To build the list of cognitive biases that may manifest while performing the fact-checking process, we started by defining three eligibility criteria to include or not a given work:

1. Is a given bias described in a peer-reviewed literature work?
2. Does the bias have a clear definition? Are its causes and domains of application explained?
3. Can we frame a fact-checking related scenario which involves the bias, eventually supported by existing literature?

The PRISMA statement emphasizes the importance of obtaining a balance between precision and recall, depending on the goals of the review. Keeping this in mind, we have defined our eligibility criteria to identify all cognitive biases with the outlined characteristics. Biases were excluded if they were not well-established, lacked a clear definition, or were not relevant to fact-checking.

Concerning information sources, more than 200 biases are listed on publicly available web pages. Specifically, Wikipedia lists 227 biases,[8] while The Decision Lab, an applied research firm, provides a list of 103 biases.[9] Other researchers have listed cognitive biases as well. For instance, Dimara, Franconeri, Plaisant, Bezerianos, and Dragicevic (2020) identifies 154 cognitive biases, and Hilbert (2012) lists 9. The search strategy involved exploring the literature retrieved by performing manual searches using each bias name as a query. The databases we utilized are Google Scholar, Scopus, PubMed, Wiley Online Library, ACL Anthology, and DBLP.

4.3. Data collection and selection processes

The complete methodology for data collection and selection that we adhere to is derived from the diagram shown in Fig. 1 and presented in Fig. 2. We consolidated the lists obtained from the information sources by removing duplicates and performing disambiguation for each bias, thus obtaining the final amount of 221 cognitive biases. Given that a standard conceptualization or classification of biases is a debated issue (Gigerenzer & Selten, 2008; Hilbert, 2012), and as our objective is to maximize the number of cognitive biases identified, we include two biases even if their differences are subtle.

To select the cognitive biases that might manifest during the fact-checking process, we analyzed each of the 221 cognitive biases found in the literature, consolidated in our list, one at a time. We focused on their definitions, causes, and domains of application, evaluating each bias according to the eligibility criteria as described in Section 4.2. The selection process has been carried out as follows: two authors of the paper (referred to as Assessor 1 and Assessor 2) individually and independently examined the full set of 221 biases, analyzing each bias definition along with a collection of practical examples for each bias. They also provided justification for the inclusion or exclusion of specific biases by citing examples of their manifestation in a fact-checking scenario. Then, Assessor 1 and Assessor 2 compared their respective lists, discussed any conflicts that arose, and reached a consensus. In order to maximize recall, they chose to include a bias in the list even if its likelihood of manifesting was relatively low. After such a step, a third author (referred to as Assessor 3) reviewed the finalized, conflict-free list of biases to ensure its comprehensiveness

[8] https://en.wikipedia.org/wiki/List_of_cognitive_biases
[9] https://thedecisionlab.com/biases/

and consistency. While the selection process inherently involves some degree of subjectivity, we believe that our implementation of discussion points, redundancy, and cross-checks establishes a robust and reliable methodology for identifying relevant cognitive biases in the context of fact-checking.

The process detailed lead to a list of 39 cognitive biases that might manifest while performing fact-checking. To the best of our knowledge, this is the first time such a list of cognitive biases will be made public. It must be noted that proposed list of cognitive biases should not be taken as final, but rather should be updated as new evidence of effects of specific cognitive biases get published in the literature.

5. List of cognitive biases

In this section, we list in alphabetical order the 39 cognitive biases that might manifest while performing fact-checking found by following the process described in Section 4; for each of them, we provide reference to literature that propose a psychological explanation for it, a short description, and we frame a situation where such bias can manifest. We also provide a fact-checking related reference to support our framing, when available. The list of 39 cognitive biases selected is presented in the following, while the full list of the 221 cognitive biases considered is reported in Appendix B.

B1. Affect Heuristic (Slovic, Finucane, Peters, & MacGregor, 2007). To often rely on emotions, rather than concrete information, when making decisions. This allows one to conclude quickly and easily, but can also distort the reasoning and lead to making suboptimal choices. This bias can manifest when the assessor likes, for example, the speaker of an information item.

B2. Anchoring Effect (Ni, Arnott, & Gao, 2019). To rely too much on an information item (typically the first one acquired) when making a decision. This bias can occur when the assessor inspects more than one source of information when assessing the truthfulness of an information item (Stubenvoll & Matthes, 2022).

B3. Attentional Bias (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007). To misperceive because of recurring thoughts. This effect may occur due to the overwhelming amount of certain topics on news media over time, for example for an assessor who is asked to evaluate the truthfulness of COVID-19 related information items (Lee & Lee, 2023).

B4. Authority Bias (also called Halo Effect) (Ries, 2006). To attribute higher accuracy to the opinion of an authority figure (unrelated to its content) and be more influenced by that opinion. This bias can manifest when the assessor is shown the speaker/organization making the information item (Javdani & Chang, 2023).

B5. Automation Bias (Cummings, 2004). To rely on automated systems which might override correct decisions made by a human assessor. This bias can occur when the assessor is presented with the outcome of an automated system that is designed to help him/her to make an informed decision on a given information item.

B6. Availability Cascade (Kuran & Sunstein, 1998). To attribute a higher plausibility to a belief just because it is public and more "available". This bias might occur when the information item presented to the assessor contains popular beliefs or popular facts (Shatz, 2020a).

B7. Availability Heuristic (Groome & Eysenck, 2016): to overestimate the likelihood of events that are recent in the memory. This bias can occur when the assessors are evaluating recent information items (Hayibor & Wasieleski, 2009).

B8. Backfire Effect (Wood & Porter, 2019). To increase one's own original belief when presented with opposed evidence. This bias, which is based on a typical human reaction, can in principle always occur in fact-checking (Swire-Thompson, DeGutis, & Lazer, 2020).

B9. Bandwagon Effect (Kiss & Simonovits, 2014). To do (or believe) things because many other people do (or believe) the same. This bias manifests for example when an assessor is asked to evaluate an information item related to recent or debated topics, for which the media coverage is high.

B10. Barnum Effect (also called Forer Effect) (Fichten & Sunerton, 1983). To fill the gaps in vague information by including personal experiences or information. This bias can in principle always occur in fact-checking (Escola-Gascon, Dagnall, Denovan, Drinkwater, & Diez-Bosch, 2023).

B11. Base Rate Fallacy (Welsh & Navarro, 2012). To focus on specific parts of information which support an information item and ignore the general information. This bias is related to the fact that the assessors are asked to report the piece of text or sources of information motivating their assessment.

B12. Belief Bias (Leighton & Sternberg, 2004). To attribute too much logical strength to someone's argument because of the validity of the conclusion. This bias is most likely to occur when the assessors are asked to evaluate factual information items (Porter & Wood, 2021).

B13. Choice-Supportive Bias (Kafaee, Marhamati, & Gharibzadeh, 2021). To remember one's own choices as better than they actually were. This bias might occur when an assessor is asked to perform a task more than one time, or when s/he is asked to revise their judgment; it might prevent assessors from revising their initially submitted score (Lind, Visentini, Mäntylä, & Del Missier, 2017).

B14. Compassion Fade (Leighton & Sternberg, 2004). To act more compassionately towards a small group of victims. This bias can occur for example when the information item to be evaluated is related to minorities, or tragic events (Thomas, Cary, Smith, Spears, & McGarty, 2018).

B15. Confirmation Bias (Nickerson, 1998). To focus on or to search for the information item which confirms prior beliefs. This bias can in principle always occur in fact-checking, for example if the assessor receives a true information that contradicts their prior beliefs, or if the assessor is asked to provide supporting evidence for their evaluation of such an information item.

B16. Conjunction Fallacy (also called Linda Problem) (Tversky & Kahneman, 1983). To assume that a conjunct event is more probable than a constituent event. Tversky and Kahneman (1983) presented this effect with a well-known example. They propose the following description: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations." and ask which of the following alternative is more probable: that (a) Linda is a bank teller (constituent event), or (b) Linda is a bank teller and is active in the feminist movement (conjunct event). Participants' intuitive methods for assessing probability often resulted in them concluding that the latter option was more likely than the former. Thus, this bias can potentially arise when an information item attributes simultaneous occurrences to a specific causal event, a pattern often found in conspiracy theories related to COVID-19 (Wabnegger, Gremsl, & Schienle, 2021).

B17. Conservatism Bias (Luo, 2014). To revise one's belief insufficiently when presented with new evidence. Note that this bias is different from B15. Confirmation Bias: the former deals with the revision of a belief, while the latter deals with new information. This bias, like B15. Confirmation Bias, can occur when the assessor is asked to provide supporting evidence for their evaluation of such an information item.

B18. Consistency Bias (Clark & Kashima, 2007). To attribute past events as resembling present behavior. This bias might occur when the assessor has evaluated an information item in the past and is asked to assess another information item coming from the same speaker and/or party.

B19. Courtesy Bias (Jones, 1963). To give a socially accepted answer to avoid offending anyone. This bias is often influenced by the assessor's personal experience and background, as well as the specific context in which they is providing their answer.

B20. Declinism (Etchells, 2015). To see the past with a positive connotation and the future with a negative one. This bias is related to the temporal part of the information items that the assessor is evaluating (Ralston, 2022).

B21. Dunning-Kruger Effect (Dunning, 2011). To overestimate oneself competence due to a lack of knowledge and skill in a certain area. This bias can occur typically to non-expert individuals, e.g., when an assessor is not trained and is overconfident about a given subject or certain topics. It is more likely to manifest with non-expert assessors in general, such as crowd workers, than expert fact-checkers like journalists.

B22. Framing Effect (Malenka, Baron, Johansen, Wahrenberger, & Ross, 1993). To draw different conclusions from logically equivalent information items based on the context, the alternatives, and the presentation method. This bias is likely to manifest, for example, when negative and positive equivalents of information items to assess refer to two counterparts of a property that have respectively negative and positive connotations, and its presentation is in terms of the share that belongs to either one or the other of these counterparts (Lindgren et al., 2022). For example, let us hypothesize that a natural disaster kills 400 out of 1000 people, thus leaving 600 alive. An information item could describe its outcome by stating that "60% of people survived the disaster" or "40% of people did not survive the disaster"; that is, by framing the fact either positively or negatively.

B23. Fundamental Attribution Error (Harvey, Town, & Yarkin, 1981). To under-emphasize situational and environmental factors for the behavior of an actor while over-emphasizing dispositional or personality factors. This bias is likely to manifest while fact-checking information items made by politically aligned speakers or news outlets, as in the case of a politician who says that people are poor because they are lazy. Furthermore, it is more likely to affect younger or older human assessors, since age differences influence its manifestation (Follett & Hess, 2002).

B24. Google Effect (Brabazon, 2006). To forget information that can be found readily online by using search engines. This bias can manifest when a worker is required to use a search engine to find evidence, and/or when they are asked to assess an information item at different time spans. For example, an assessor forgetting part of the information item right after reading it, because they know that it is easily retrievable again if needed by querying a search engine (Lurie & Mustafaraj, 2018).

B25. Hindsight Bias (also called "I-knew-it-all-along" Effect) (Roese & Vohs, 2012). To see past events as being predictable at the time those events happened. Since it may cause distortions of memories of what was known before an event, this bias may manifest when an assessor is required to evaluate an event after some time or when is asked to evaluate the same information item multiple times at different time spans (Hom, 2022).

B26. Hostile Attribution Bias (Pornari & Wood, 2010). To interpret someone's behavior as hostile even if it is not. This bias can occur for assessors who have experienced discrimination from authority figures or the dominant social group. For example, a speaker from an under-represented ethnicity remarks oneself that they perceive something as offensive, thus fueling hostility (Bushman, 2016).

B27. Illusion of Validity (Einhorn & Hogarth, 1978). To overestimate someone's judgment when the available information is consistent. This bias can occur for example when an assessor works with a set of previously true information items from a specific person and predicts that the subsequent set of information items will have the same outcome from the same person.

B28. Illusory Correlation (Hamilton & Gifford, 1976). To perceive the correlation between non-correlated events. This bias can manifest when an assessor works on multiple information items in a single task and may perceive nonexistent patterns between the items.

B29. Illusory Truth Effect (Newman, Schwarz, & Ly, 2020). To perceive an information item as true if it is easier to process or it has been stated multiple times. This bias can manifest for example when using straightforward or naive gold questions in a task to check for malicious assessors (Brashier, Eliseev, & Marsh, 2020).

B30. Ingroup Bias (Mullen, Brown, & Smith, 1992). To favor people belonging to one's own group. This bias can manifest for example when the assessors are required to work on information items related to their own political party, city, etc (Shin & Thorson, 2017).

B31. Just-World Hypothesis (Lerner & Miller, 1978). To believe that the world is just. This bias can happen for example when the assessor is working with information items related to major political institutions, as people tend to assign to them higher scores in belief (Rubin & Peplau, 1975).

B32. Optimism Bias (also called Optimistic Bias) (Sharot, 2011). To be over-optimistic, underestimating the probability of undesirable outcomes and overestimating favorable and pleasing outcomes. This bias can occur for example when the information item provides some kind of assessment of the risk of an event manifesting, such as the likelihood of getting infected by COVID-19 (Druică, Musso, & Ianole-Călin, 2020).

B33. Ostrich Effect (also called Ostrich Problem) (Karlsson, Loewenstein, & Seppi, 2009). To avoid potentially negative but useful information, such as feedback on progress, to avoid psychological discomfort. This bias can occur when an assessor avoids evaluating, explicitly or not, an information item which has a negative connotation according to their own beliefs.

B34. Outcome Bias (Baron & Hershey, 1988). To judge a decision by its eventual outcome instead on the basis of the quality of the decision at the time it was made. This bias can manifest when the information item under consideration is related to a past event (Robson, 2019).

B35. Overconfidence Effect (Dunning, Griffin, Milojkovic, & Ross, 1990). To be too confident in one's own answers. This effect can manifest when the assessor is an expert in the field, as for example an expert journalist who performs fact-checking related to their writing or a medical specialist who assesses health-related information items.

B36. Proportionality Bias (also called Major Event/Major Cause Heuristic) (Leman & Cinnirella, 2007). To assume that big events have big causes. This innate human tendency can also explain why some individuals accept conspiracy theories. This bias can occur when the factual information being assessed deals with the causes and effects of a particular event (Stall & Petrocelli, 2023).

B37. Salience Bias (Mullen et al., 1992). To focus on items that are more prominent or emotionally striking and ignore those that are unremarkable, even though this difference is irrelevant by objective standards. For example, an information item detailing the numerous deaths of infants will receive more attention than an information item detailing a less emotionally striking fact. If those two facts are presented in the same information item, the score assessed for the prominent fact might drive the overall assessment of the whole information item.

B38. Stereotypical Bias (Heilman, 2012). To discriminate against a personal trait (e.g., gender). Like B30. Ingroup Bias, this bias can happen when the assessor, especially a crowd worker, identifies themselves with the group related to the information item they is assessing.

B39. Telescoping Effect (Thompson, Skowronski, & Lee, 1988). To displace recent events backward in time and remote events forward in time, so that recent events appear more remote, and remote events, more recent. This bias might occur when the information item presented contains temporal references.

6. Categorization of cognitive biases

Considering at the same time all the 39 biases that might manifest while performing fact-checking can be challenging for researchers and practitioners that aim studying their manifestation and/or impact in fact-checking settings. Thus, providing a second level of aggregation might support laying out the problem and facilitate further analysis, for instance by considering the type of fact-checking related task. To this end, we further categorize the 39 biases from a task-based perspective, utilizing the classification scheme proposed by Dimara et al. (2020) and employed to address cognitive biases specifically affecting information visualization tasks. This scheme allows for aggregating our initial list based on the psychological explanations of why biases might occur in a fact-checking related context, as reported in Section 5.

Dimara et al. scheme for classifying cognitive biases works as follows. They first identify user tasks that might involve cognitive biases, thus generating a set of 7 task categories (Dimara et al., 2020, Section 3.4) using open card-sorting analysis (Wood & Wood, 2008). The user task categories found are summarized in Table 3. Then, they focus on the relevance of cognitive biases with information visualization aspects. The initial classification of cognitive biases according to the user task included a fairly large number of biases for each category. Thus, they further refined the overall scheme by proposing a set of 5 sub-categories called flavors (Dimara et al., 2020, Section 3.4) that focus on other types of similarities across cognitive biases, related with how each bias affects human cognition. Such flavors are summarized in Table 4.

We categorize our set of 39 biases that might manifest in the fact-checking activity by assessing which type of task they might affect and how they influence human cognition, that is, by assigning them with a given task/flavor combination according to the scheme by Dimara et al. (2020, Table 2). The categorization process involved evaluating each bias identified in our initial list against the seven task categories and five flavors. This was done by first determining the most likely fact-checking task each bias could influence (Table 3). Subsequently, we analyzed how each bias affects human cognition by aligning them with one of the flavors (Table 4). This dual-level categorization allowed for a detailed understanding of how each bias could potentially manifest in various aspects of fact-checking. Among our list of 39 cognitive biases, there are 35 biases that both we and Dimara et al. consider (although in two different contexts); for them the two classifications agree. Furthermore, there are four biases that Dimara et al. did not consider in their classification. Such biases are: B18. Consistency Bias, B19. Courtesy Bias, B36. Proportionality Bias, and B37. Salience Bias. We conducted an in-depth analysis to appropriately assign these biases to both a task category and a flavor. This involved assessing the nature and implications of each bias and determining the most relevant task and flavor based on their characteristics and impact on cognitive processes during fact-checking.

Table 3
Types of user tasks that may involve cognitive biases, as proposed by Dimara et al. (2020).

Task	Description
Causal Attribution	Tasks involving an assessment of causality.
Decision	Tasks involving the selection of one over several alternative options.
Estimation	Tasks where people are asked to assess the value of a quantity.
Hypothesis Assessment	Tasks involving an investigation of whether one or more hypotheses are true or false.
Opinion Reporting	Tasks where people are asked to answer questions regarding their beliefs or opinions on political, moral, or social issues.
Recall	Tasks where people are asked to recall or recognize previous material.
Other	Tasks which are not included in one of the previous categories.

Table 4
Phenomena that affect human cognition, as proposed by Dimara et al. (2020).

Flavor	Description
Association	Cognition is biased by associative connections between information items.
Baseline	Cognition is biased by comparison with (what is perceived as) a baseline.
Inertia	Cognition is biased by the prospect of changing the current state.
Outcome	Cognition is biased by how well something fits an expected or desired outcome.
Self-Perspective	Cognition is biased by a self-oriented viewpoint.

The task/flavor classification described in Table 5 provides a structured and detailed approach to understanding the multifaceted ways in which cognitive biases can influence the fact-checking process. We acknowledge that differently from the selection of the 39 biases, where a structured PRISMA-based approach was possible, this classification is necessarily more subjective, as it is achieved by agreement between evaluators. By mapping each bias to specific fact-checking tasks and cognitive influences, we aim to offer a comprehensive framework that aids researchers and practitioners in identifying and addressing potential biases in their work.

To provide an interpretation of the categorization scheme based on tasks and flavors in a fact-checking context, let us make an example of a Decision task as defined by Dimara et al. (2020) (see Table 3). As these types of tasks involve selecting one option from several alternatives, let us consider a scenario where an assessor is asked to determine which information item is more truthful among two different alternatives. Let us further hypothesize that the two information items are made by politicians, with one belonging to the governing party. In such a case, one may argue that since the trustworthiness of the speaker might be linked to B4. Authority Bias, the assessor might believe that being part of the governing party implies higher trustworthiness for the speaker. Moreover, since reasoning (or, as Dimara et al. call it, cognition) is biased by an associative connection between the two pieces of information, we conclude that the underlying flavor is Association.

7. List of countermeasures

The literature allows us to specify 11 countermeasures that can be employed in a fact-checking context to help prevent manifesting the cognitive biases outlined in Table 5. We detail each countermeasure in the following (C1–C11).

To select the countermeasures, we proceed as follows. First, we inspect the literature to identify works aiming at addressing specific biases, and second, we select the proposed countermeasures from those works that can be applied when performing the fact-checking activity. Note that this approach only details how to remove individual biases, but it must be noted that the removal of one bias as a result of the application of a countermeasure might result in a manifestation of another one. For instance, Park et al. (2021) show that often there are unexpected biases that arise in a fact-checking scenario, thus there might not exist a systematic way to safely remove all the possible sources of bias altogether. Hence, researchers and practitioners should aim at finding a good compromise between the possibility of bias manifestation and the specific experimental setting. The 11 identified countermeasures are listed in the following, in alphabetical order; for each countermeasure we cite the relevant literature examined.

Table 5
Categorization of cognitive biases, adapted from Dimara et al. (2020).

	Association	Baseline	Inertia	Outcome	Self-Perspective
Causal Attribution	–	–	–	B26. Hostile Attribution Bias B31. Just-World Hypothesis	B23. Fundamental Attribution Error B30. Ingroup Bias
Decision	B4. Authority Bias B5. Automation Bias B22. Framing Effect	–	–	–	–
Estimation	B7. Availability Heuristic B16. Conjunction Fallacy	B2. Anchoring Effect B11. Base Rate Fallacy B14. Compassion Fade B21. Dunning-Kruger Effect B35. Overconfidence Effect	B17. Conservatism Bias	B27. Illusion of Validity B34. Outcome Bias	B32. Optimism Bias B37. Salience Bias
Hypothesis Assessment	B6. Availability Cascade B29. Illusory Truth Effect	–	–	B10. Barnum Effect B12. Belief Bias B15. Confirmation Bias B28. Illusory Correlation	–
Opinion Reporting	–	B36. Proportionality Bias	B8. Backfire Effect	B9. Bandwagon Effect B38. Stereotypical Bias	B19. Courtesy Bias
Recall	B24. Google Effect B39. Telescoping Effect	–	B18. Consistency Bias	B13. Choice-Supportive Bias B20. Declinism B25. Hindsight Bias	–
Other	B3. Attentional Bias	–	–	B33. Ostrich Effect	–

C1. Custom search engine. Researchers and practitioners should be extremely careful with the system supplied to the assessors to help them retrieving some kind of supporting evidence, since such a system can be biased (Diaz, 2008; Mowshowitz & Kawaguchi, 2005; Otterbacher, Checco, Demartini, & Clough, 2018; Wilkie & Azzopardi, 2014). Researchers should employ a custom and controllable search engine when asking the assessors to evaluate an information item. The assessors might be influenced by the score assigned to the news by a news agency or an online website for the very same information item. Thus, the researcher may tune the search engine parameters to limit the bias that each assessor encounters during a fact-checking activity due to the result source.

C2. Inform assessors. Researchers should always inform assessors about the presence of any kind of automatic (e.g., AI-based) system designed to provide support during the assessment activity, for example by asking them for confirmation or rejection while using such systems, thus limiting B5. Automation Bias (Goddard, Roudsari, & Wyatt, 2011; Kupfer et al., 2023). This includes, for example, the presence of a search engine helping them finding some kind of evidence (Draws et al., 2022; Roitero, Soprano, Fan, et al., 2020; Roitero, Soprano, Portelli, et al., 2020; Soprano et al., 2021).

C3. Discussion. Researchers should allow a synchronous discussion among assessors when possible. In fact, when evaluating the truthfulness of an information item each individual is more prone to accept information items that are consistent with their set of beliefs (La Barbera et al., 2020; Lewandowsky et al., 2012; Roitero, Soprano, Fan, et al., 2020). Reimer, Reimer, and Czienskowski (2010) and Szpara and Wylie (2005), indeed, proved the effectiveness of conducting a synchronous discussion between different assessors to reduce their own bias. Pitts, Coles, Thomas, and Smith (2002) and Zheng, Cui, Li, and Huang (2018) show how discussion among assessors improves the overall assessment quality.

C4. Engagement. It is important to put the assessors in a good mood when performing a fact-checking task. Cheng and Wu (2010) show that engaged assessors are less likely to experience both:

- B22. Framing Effect.

  - B28. Illusory Correlation.

  Moreover, Furnham and Boo (2011) show that if assessors are engaged they are less likely to experience:

  - B2. Anchoring Effect.
  - B33. Ostrich Effect.

C5. Instructions. Another important aspect to consider consists of formulating an adequate set of instructions. Gillier, Chaffois, Belkhouja, Roth, and Bayus (2018) have shown that a set of instructions helps assessors in coming up with new ideas when performing a crowdsourcing task. Even though Gadiraju, Yang, and Bozzon (2017) explain that assessors can perform a task even if they have a sub-optimal understanding of the work requested, task instructions clarity should be taken into account. Furthermore, the assessors should be encouraged explicitly to be skeptical about the information that they are evaluating (Lewandowsky et al., 2012). Indeed, Ecker, Lewandowsky, and Tang (2010) and Schul (1993) prove that pre-exposure warning (i.e., telling explicitly a person that they could be exposed to something) reduces the overall impact on the person itself. Thus, showing a set of assessment instructions can be seen as a pre-exposure warning against the impact of misinformation on the assessor.

C6. Require evidence. Requiring the assessors to provide supporting evidence for their judgments is another effective countermeasure with several advantages. It encourages the assessor to focus on verifiable facts. Lewandowsky et al. (2012) explain that such a countermeasure increases the perceived familiarity with the information item, reinforcing the assessor perceived trustworthiness of the information item itself. They also show that reporting a small set of facts as evidence has the effect of discouraging possible criticisms by other assessors, thus reinforcing the assessment provided. Jerit (2008) observes such a phenomenon in public debates. Furthermore, asking the assessors to come up with arguments to support their assessment has proven to reduce:

  - B2. Anchoring Effect, as shown by Mussweiler, Strack, and Pfeiffer (2000).
  - B11. Base Rate Fallacy, as shown by Kahneman and Tversky (1973).
  - B22. Framing Effect, as shown by Cheng and Wu (2010), Kim, Goldstein, Hasher, and Zacks (2005).
  - B27. Illusion of Validity, as shown by Kahneman and Tversky (1973).
  - B28. Illusory Correlation, as shown by Matute, Yarritu, and Vadillo (2011).

  However, requesting for evidence may be a source of bias itself. Luo (2014) and Wood and Porter (2019) show, indeed, that such a request can lead to the manifestation of, respectively:

  - B8. Backfire Effect.
  - B17. Conservatism Bias.

  Thus, the requester of the fact-checking activity should address this matter carefully.

C7. Randomized or constrained experimental design. Using a randomized or constrained experimental design is helpful in reducing biases. Indeed, different assessors should evaluate different information items. Moreover, each set of items should be evaluated according to a different order, and the assignment of an information item to a given assessor should be such that the item overlap between every two assessors is minimum. If such a constraint cannot be satisfied, a randomization process should minimize the chances of overlap between items and assessors (Ceschia et al., 2022; Hettiachchi, Kostakos, & Goncalves, 2022).

C8. Redundancy and diversity. Redundancy should be employed when asking more than one assessor to fact-check a set of information items. Indeed, the same information item should be evaluated by different assessors. Each item can thus be characterized by a final score, that should be computed by aggregating the individual scores provided by each assessor. In this way, the individual bias of each assessor is mitigated by the remaining assessors. If the population of assessors is diverse enough, one can ideally expect less bias from the fact-checking activity. The population of assessors should thus be as variegated as possible, in terms of both background and experience (Difallah, Filatova, & Ipeirotis, 2018).

C9. Revision. Asking the assessors to revise and/or double-check their answers or even provide them with alternative labels is a useful countermeasure to reduce many biases. In more detail, Cheng and Wu (2010), Kahneman (2011), Kahneman and Tversky (1973), Kim et al. (2005), Mussweiler et al. (2000), and Shatz (2020b) show that assessment revision helps reducing:

  - B2. Anchoring Effect.
  - B7. Availability Heuristic.
  - B9. Bandwagon Effect.
  - B11. Base Rate Fallacy.
  - B22. Framing Effect.

  Furthermore, Bollinger, Leslie, and Sorensen (2011), Cooper et al. (2014), Hettiachchi, Schaekermann, McKinney, and Lease (2021), and Mussweiler et al. (2000) show that providing feedback to assessors while performing a given task is useful to reduce biases such as:

  - B2. Anchoring Effect.
  - B3. Attentional Bias.

  - B37. Salience Bias.

C10. Time. Researchers should be careful when setting the time available for each assessor to fact-check a given information item. An adequate amount of time should be left to the assessor. There are advantages and disadvantages of granting the assessor with a small or large amount of time. For instance, one may assume that providing the assessor with more time will encourage careful consideration of the decision, thus helping to avoid the B2. Anchoring Effect. However, Furnham and Boo (2011) show that overthinking might actually increase such a bias. On the other hand, Shatz (2020b) shows that assessors left with an adequate amount of time experienced a reduction of the B9. Bandwagon Effect.

C11. Training. Dugan (1988), Kazdin (1977), Lievens (2001), Pell, Homer, and Roberts (2008), and Szpara and Wylie (2005) show that training an assessor increases accuracy and reduces the chances for bias to manifest. Thus, assessors training is a useful countermeasure against biases within any context.

8. Towards a bias-aware assessment pipeline for fact-checking

In Section 7 we presented 11 countermeasures to reduce the risk of manifestation of the 39 cognitive biases listed in Section 5 and further categorized in Section 6. We can thus leverage them to propose the constituting elements of a fact-checking pipeline that minimizes such a risk, outlined in Table 6. The table shows, for each part of the overall fact-checking activity, the corresponding countermeasures along with a brief recap and the biases involved.

More specifically, the first column of the table details the task phase where the specific countermeasure can be applied: before the task (i.e., pre-task), when the assessor is performing the task (i.e., during the task), or after the task, when the assessment has been made (i.e., post-task); furthermore, we list a set of countermeasures that are not bounded to a specific task purpose (i.e., general purpose). The remaining columns detail the countermeasures that can be adopted within the corresponding phase and, for each of them, a brief description together with the set of biases that they reduce.

Let us make it more clear by providing an example; the first row of the table deals with the adoption of the countermeasure C7. Randomized or constrained experimental design, i.e., to randomize the process that assigns assessors and information items to enforce diversity and randomness in the pairing. This is a general purpose countermeasure since it can be applied in different task phases (e.g., when designing the task offline or dynamically when a new assessor is assigned to a new information item). It allows for the mitigation or removal of the B2. Anchoring Effect, as the assessor is less likely to rely on a specific information item, given that they inspect more than one with different characteristics. Additionally, it helps address the B9. Bandwagon Effect, as the assessor is less likely to be presented with a set of items all related to their personal beliefs or to debated topics with high coverage.

In light of the detailed review of current fact-checking practices by prominent organizations like PolitiFact, RMIT ABC Fact Check, and FactCheck.org presented in Section 2.1, we can now discuss how our proposal of a bias-aware pipeline can be considered as an enhancement to these existing practices. The key distinction lies in our approach's systematic integration of cognitive bias countermeasures at various stages of the fact-checking process, which is not explicitly addressed in the conventional practices followed by organizations.

Our proposed pipeline, as outlined in Table 6, includes specific countermeasures targeting known cognitive biases. For instance, PolitiFact's process involves consensus among editors and reporters for the final rating, which can be influenced by biases like B2. Anchoring Effect or B9. Bandwagon Effect. Our pipeline suggests measures like randomized pairing of assessors and information items, and synchronous discussion between assessors, to mitigate these biases. Similarly, RMIT ABC Fact Check's process, involving a collaborative review where a team decides the final verdict, could benefit from our proposed countermeasures like requiring evidence to support assessments and revising the assessments to encourage a systematic evaluation of the evidences that support the assessments, to counter biases such as B22. Framing Effect or B28. Illusory Correlation. FactCheck.org's approach, emphasizing evidence retrieval and a review team, aligns with our recommendation of using a custom search engine and informing assessors about AI-based support systems to reduce the impact of biases like B5. Automation Bias.

In summary, while the existing practices of these organizations adhere to high standards of transparency, accuracy, and thoroughness, the integration of our bias-aware pipeline could further enhance the objectivity and reliability of the fact-checking process by explicitly addressing and mitigating the influence of cognitive biases. This not only would strengthen the trustworthiness of the fact-checking results, but would also align with the evolving needs of a complex information landscape where biases can significantly impact the interpretation and verification of information. In the future, we plan to engage directly with organizations like PolitiFact, RMIT ABC Fact Check, and FactCheck.org to discuss the practical testing and implementation of such a bias-aware pipeline in their fact-checking processes.

9. Limitations

While our research provides valuable insights into the set of cognitive biases affecting human assessors during the fact-checking process, there are some limitations of our work that should be acknowledged.

First, there is subjectivity involved both in the processes of identification and classification of the 221 cognitive biases, which are based on the authors' interpretation and understanding of the individual biases, and in the creation of the fact-checking scenarios that involve each of these cognitive biases. Despite our efforts to provide reasonable and grounded examples, for 23 out of 39 biases we have been able to support our scenario formulation by relying on the literature, but for the remaining 16 biases we have not been able to find a fact-checking related reference. Thus, in these cases, the inherent subjectivity of our formulation must be acknowledged.

Table 6
Constituting elements of a bias-aware assessment pipeline.

Task Phase	Countermeasure	Brief Description	Biases Involved
General Purpose	C7. Randomized or constrained experimental design	Employ a randomization process when pairing assessors and information items	Bias in General
	C8. Redundancy and diversity	Use more than one assessor for each information item, and a variegated pool of assessors	Bias in General
	C10. Time	Allocate an adequate amount of time for the assessors to perform the task	B2. Anchoring Effect B9. Bandwagon Effect
Pre-Task	C4. Engagement	Put the assessors in a good mood and keep them engaged	B2. Anchoring Effect B22. Framing Effect B28. Illusory Correlation B33. Ostrich Effect
	C5. Instructions	Prepare a clear set of instructions to the assessors before the task	B21. Dunning-Kruger Effect B35. Overconfidence Effect
	C11. Training	Train the Assessors before the task	Bias in General
During the Task	C1. Custom search engine	Deploy a custom search engine	Bias in General
	C2. Inform assessors	Inform the assessors about AI-based support systems	B5. Automation Bias
	C3. Discussion	Synchronous discussion between assessors	Bias in General
	C6. Require evidence	Ask the assessors to provide supporting evidence	B2. Anchoring Effect B8. Backfire Effect B11. Base Rate Fallacy B17. Conservatism Bias B22. Framing Effect B27. Illusion of Validity B28. Illusory Correlation
	C9. Revision	Ask the assessors to revise the assessments	B2. Anchoring Effect B3. Attentional Bias B7. Availability Heuristic B9. Bandwagon Effect B11. Base Rate Fallacy B22. Framing Effect B37. Salience Bias
Post-Task	C8. Redundancy and diversity	Aggregate the final scores	Bias in General

Given that these processes are to some extent subjective, different researchers and practitioners might identify and classify these biases differently. Although our PRISMA-based methodology aimed to provide a comprehensive list of cognitive biases affecting fact-checking, it is possible that some biases were overlooked or not considered due to the vast amount of literature on cognitive biases. Thus, our proposed list may not be exhaustive or complete.

Then, it is important to note that our findings may not be generalizable to all possible fact-checking contexts or human populations, as cognitive biases may manifest differently depending on the specific context and individual differences.

Regarding the set of countermeasures presented, it should be noted that it is difficult to ascertain the extent to which the countermeasures are effective and general. Their effectiveness could be influenced by various factors, such as the specific context, individual differences, and the nature of the misinformation.

We also remark that in our research we have considered each cognitive bias as independent. However, biases may interact with each other in complex ways, potentially magnifying or attenuating their effects on the fact-checking process. Future research should investigate the possible interactions between cognitive biases and their cumulative impact on the ability of human assessors to accurately evaluate information.

It should be also noted that this study is primarily based on an analysis of the literature; as such, we did not conduct any empirical tests to validate the potential impact of the identified cognitive biases on the fact-checking process involving human assessors. Future research should consider experimental activities to investigate the actual effects of these biases on assessors' performance and the effectiveness of the proposed countermeasures.

Finally, it should be noted that the cognitive biases we identified and discussed in this paper are based on the current state of knowledge. As the field of cognitive psychology continues to evolve, new biases may be discovered or existing ones may be refined or redefined. Thus, our findings and analyses should be periodically updated to reflect the most current understanding of cognitive biases and their potential impact on the fact-checking process.

10. Conclusion and future work

In this review, we discuss the problem of cognitive biases that may manifest while performing fact-checking tasks. Our study summarizes the literature and leverages the "PRISMA 2020 Statement" to systematically identify and categorize these biases, ensuring a rigorous and comprehensive approach to the problem. To our knowledge, this is the first attempt to comprehensively study – and handle – the cognitive biases that may be present at the different stages of the fact-checking workflow.

We have identified and detailed a subset of 39 out of 221 cognitive biases that are relevant to fact-checking, discussing each of them in detail, and formulating examples of their manifestation in real-world scenarios (Section 5). Furthermore, we propose a classification of such biases (Section 6) and we provide a list of countermeasures to limit their impact (Section 7). Finally, we outline the constituting elements of a bias-aware fact-checking pipeline and we assign our countermeasures to each block (Section 8).

This paper, and in particular the proposed bias-aware assessment pipeline for fact-checking, has many implications. Researchers and practitioners are allowed to fully understand which cognitive biases can manifest in a fact-checking task. In particular, the findings presented in this work may help human expert fact-checkers in revising their processes (Ceci & Williams, 2020) as well as help artificial intelligence developers in building models that are robust against biased data and result in fair decisions. A step in this direction is Table 6, which presents possible actions towards the mitigation of the identified risks in producing biased fact-checking decisions. For example, filtering the information available to human assessors to avoid them being biased by evidence they should not consider during their judgment process (see C1. Custom search engine), or allow for an open discussion to highlight possible extreme individual views (see C3. Discussion). Instructions given to assessors are another possible source of bias and thus making sure they are presented as intended (see C5. Instructions) is a way to avoid wrong priming and bias.

This systematic characterization of cognitive biases during the fact-checking process sets the basis for future work. Researchers and practitioners can use the set of identified cognitive biases and design ad hoc experiments to further investigate how to manage and mitigate the effects of these biases. In future studies, we propose to integrate quantitative methods with our current qualitative approach. These could involve a bibliometric analysis to identify the most influential papers within our body of cited literature. Such an analysis can offer insights into the prevalence and impact of specific cognitive biases in shaping human judgments in the field of fact-checking. In addition, statistical tools and techniques could be employed to evaluate the extent of these biases' influence. By doing so, we aim to create a more comprehensive understanding of the interplay between cognitive biases and fact-checking processes. For instance, we have seen in both Section 5 and Table 6 that asking the assessors to revise their judgment (i.e., adopting C9. Revision) might help in mitigating B2. Anchoring Effect. Thus, researchers and practitioners can set up a between-subject experiment for truthfulness classification where assessors are divided into two disjoint sets; while both set of assessors are presented with an initial information item before each assessment (i.e., the anchor point), the former set of assessors is asked to revise their truthfulness judgment before submitting it, while the latter is not. Researchers can then measure the data gathered for the two sets of annotators and empirically verify the amount of B2. Anchoring Effect in the annotations; depending on the outcome of the analysis, practitioners might decide to use the less biased experimental setting to collect the final set of judgments.

Summarizing, our review investigates the role of cognitive biases in the fact-checking process, offering valuable insights and practical guidance for researchers and practitioners in the field. By identifying and addressing these biases, we can contribute to more accurate and reliable information assessment and ultimately combat the spread of misinformation and disinformation in today's world, and use this work which serves as a reference to build methodologies to perform a more sound, robust, aware, and less biased fact-checking at scale.

CRediT authorship contribution statement

Michael Soprano: Conceptualization, Data curation, Investigation, Methodology, Resources, Writing – original draft, Writing – review & editing. **Kevin Roitero:** Conceptualization, Data curation, Investigation, Methodology, Resources, Writing – original draft, Writing – review & editing. **David La Barbera:** Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Davide Ceolin:** Methodology, Writing – review & editing. **Damiano Spina:** Conceptualization, Writing – review & editing. **Gianluca Demartini:** Conceptualization, Writing – review & editing. **Stefano Mizzaro:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Data availability

All the data involved have been provided in the appendices.

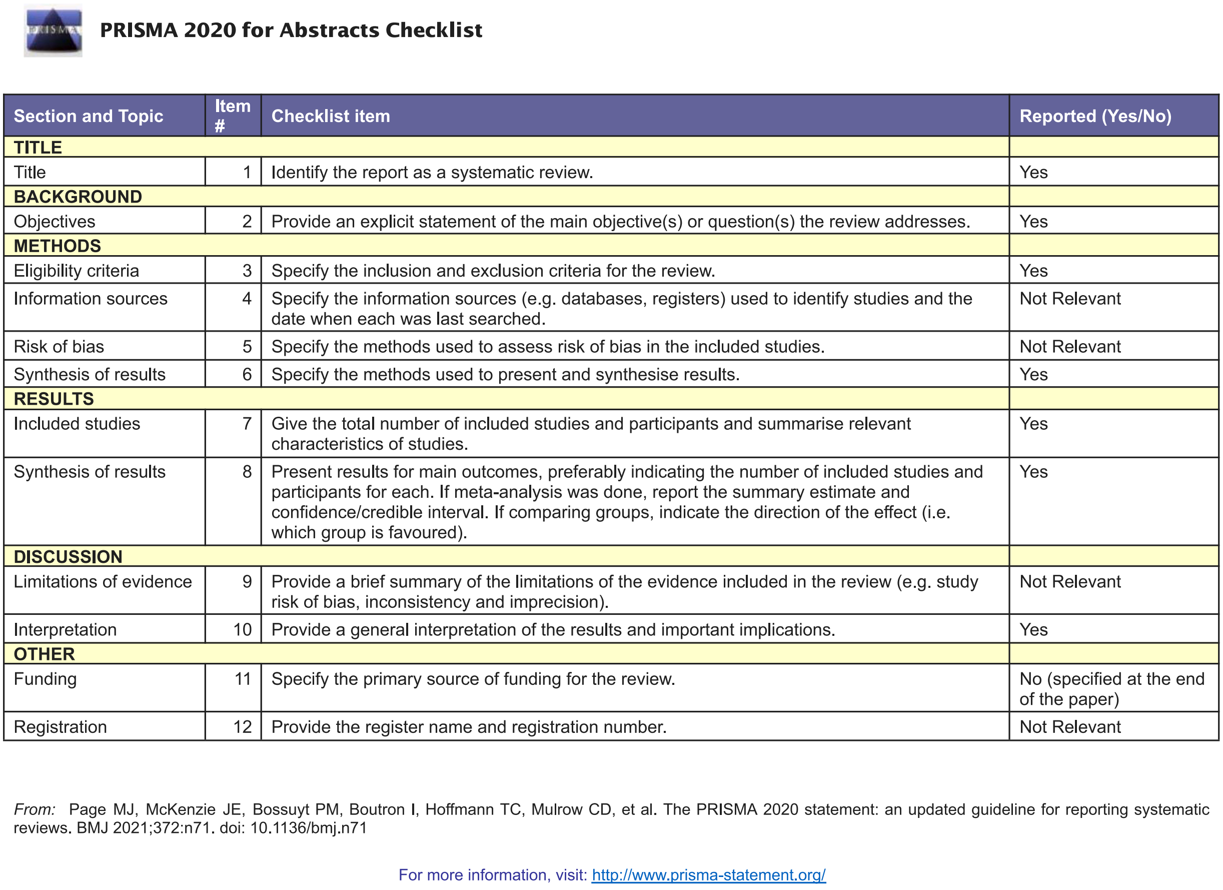
Disclosure and Acknowledgments

This research is partially supported by the Australian Research Council (DE200100064, CE200100005, IC200100022). Damiano Spina is the recipient of an Australian Research Council (ARC) DECRA Research Fellowship (DE200100064), and Associate Investigator of the ARC Centre of Excellence for Automated Decision-Making and Society (Australia) (CE200100005), and a research collaborator of RMIT FactLab. Gianluca Demartini is a Chief Investigator of the ARC Training Centre for Information Resilience (Australia) (IC200100022).

This research is also supported by the Next Generation EU PRIN 2022 project "20227F2ZN3 MoT—The Measure of Truth: An Evaluation-Centered Machine–Human Hybrid Framework for Assessing Information Truthfulness" and by the Strategic Plan of the University of Udine-Interdepartment Project on Artificial Intelligence (2020–25) (Italy). Furthermore, this publication has also been partly supported by the Netherlands eScience Center project "The Eye of the Beholder" (project nr. 027.020.G15) (Netherlands) and it is part of the AI, Media & Democracy Lab (Dutch Research Council project number: NWA.1332.20.009) (Netherlands). For more information about the lab and its further activities, visit https://www.aim4dem.nl/. Any opinions, findings, and conclusions expressed in this article are those of the authors and do not necessarily reflect those of the sponsors.

Appendix A. Prisma checklists

This appendix reports the whole checklists of the "PRISMA 2020 Statement" used to conduct the review about cognitive biases that might manifest in a fact-checking context described in Section 4. Table 2 summarizes the "Item #" and "Section and Topic" columns of the main checklist.

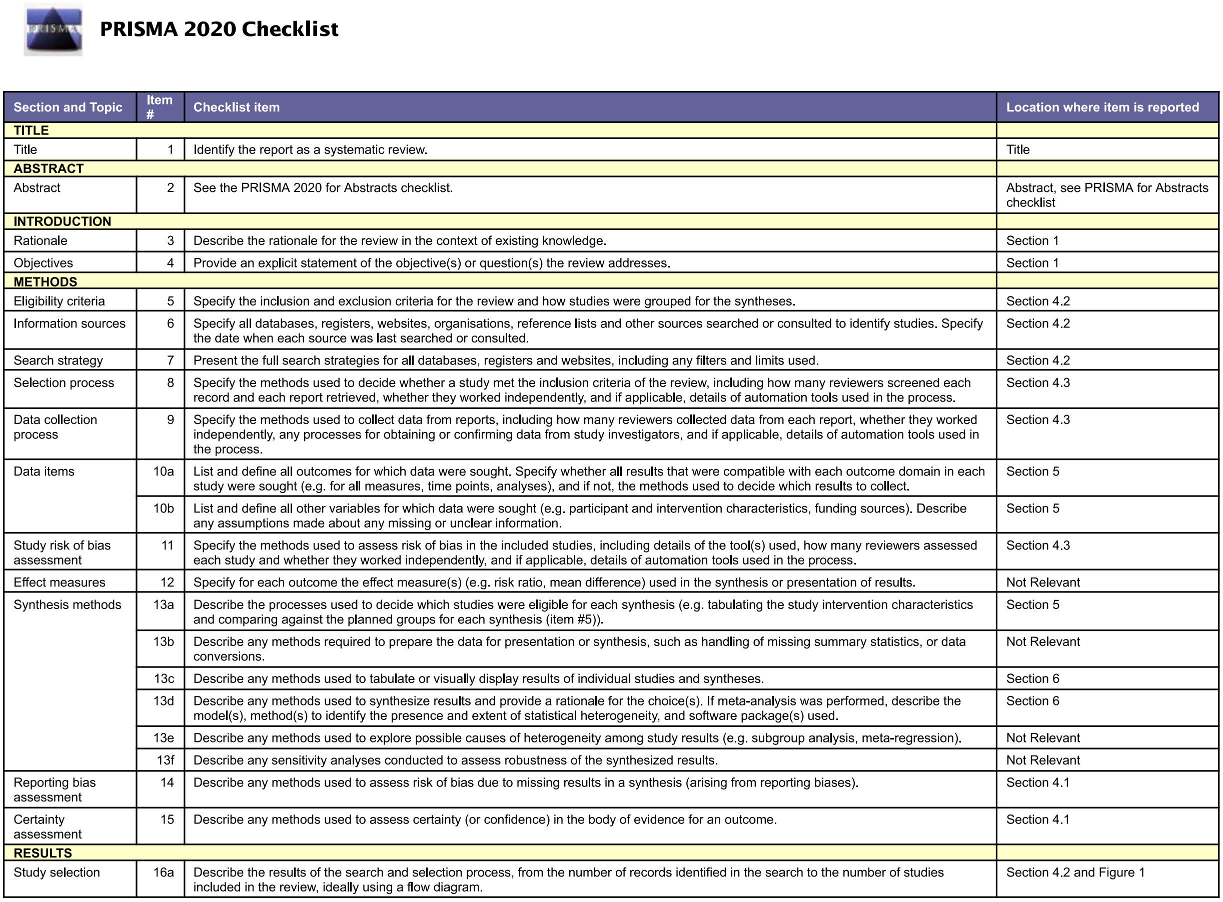


PRISMA 2020 for Abstracts Checklist

Section and Topic	Item #	Checklist item	Reported (Yes/No)
TITLE			
Title	1	Identify the report as a systematic review.	Yes
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Yes
METHODS			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	Yes
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Not Relevant
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	Not Relevant
Synthesis of results	6	Specify the methods used to present and synthesise results.	Yes
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Yes
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Yes
DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Not Relevant
Interpretation	10	Provide a general interpretation of the results and important implications.	Yes
OTHER			
Funding	11	Specify the primary source of funding for the review.	No (specified at the end of the paper)
Registration	12	Provide the register name and registration number.	Not Relevant

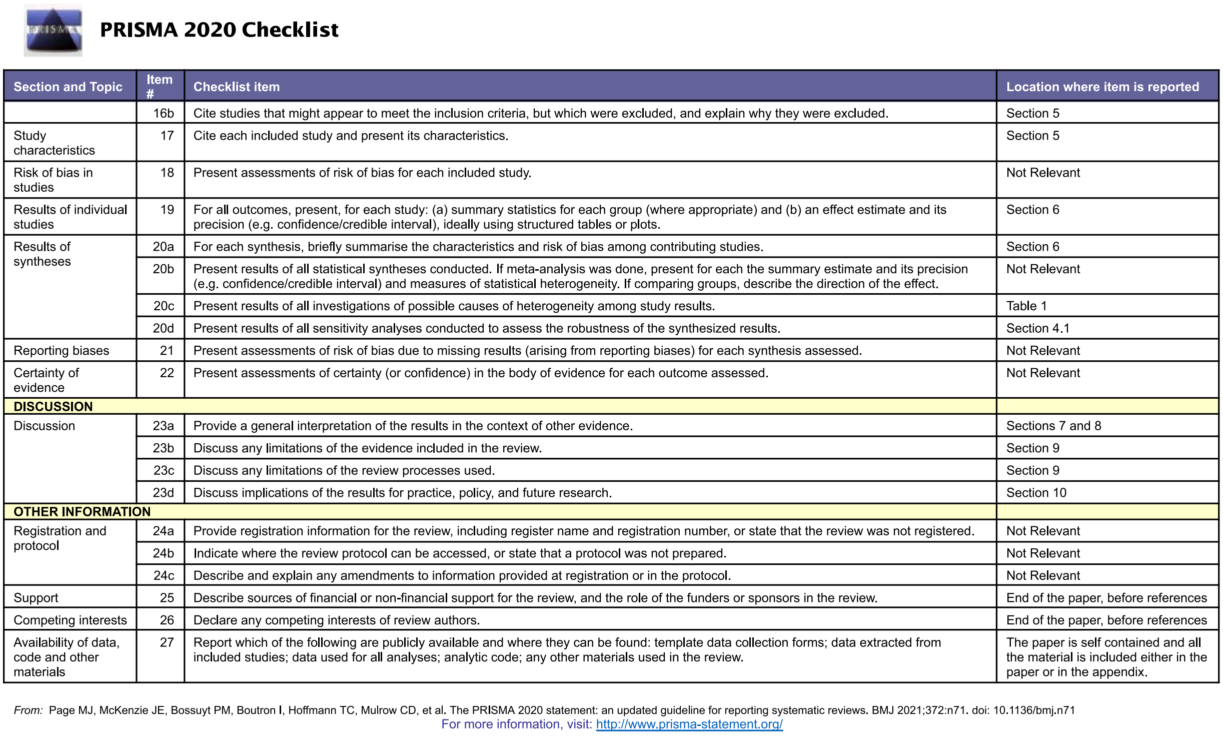
From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract, see PRISMA for Abstracts checklist
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Section 1
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Section 1
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Section 4.2
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Section 4.2
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Section 4.2
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Section 4.3
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Section 4.3
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Section 5
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Section 5
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Section 4.3
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Not Relevant
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Section 5
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Not Relevant
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Section 6
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Section 6
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Not Relevant
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Not Relevant
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Section 4.1
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Section 4.1
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Section 4.2 and Figure 1



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Section 5
Study characteristics	17	Cite each included study and present its characteristics.	Section 5
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Not Relevant
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Section 6
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Section 6
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Not Relevant
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Table 1
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Section 4.1
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not Relevant
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Not Relevant
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Sections 7 and 8
	23b	Discuss any limitations of the evidence included in the review.	Section 9
	23c	Discuss any limitations of the review processes used.	Section 9
	23d	Discuss implications of the results for practice, policy, and future research.	Section 10
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Not Relevant
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Not Relevant
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not Relevant
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	End of the paper, before references
Competing interests	26	Declare any competing interests of review authors.	End of the paper, before references
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	The paper is self contained and all the material is included either in the paper or in the appendix.

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71
For more information, visit: http://www.prisma-statement.org/

Appendix B. List of 221 cognitive biases

This appendix reports the full list 221 of cognitive biases found in the literature using the methodology described in Section 4. The 39 cognitive biases that might manifest in a fact-checking context, described in Section 5, have been derived from this list.

1. Action Bias
2. Actor-observer Bias
3. Additive Bias
4. Agent Detection Bias
5. Affect Heuristic
6. Ambiguity Effect
7. Anchoring Effect
8. Anthropocentric Thinking
9. Anthropomorphism
10. Apophenia
11. Association Fallacy
12. Assumed Similarity Bias
13. Attentional Bias
14. Attribute Substitution
15. Attribution Bias
16. Authority Bias
17. Automation Bias
18. Availability Bias
19. Availability Cascade
20. Availability Heuristic
21. Backfire Effect
22. Bandwagon Effect
23. Barnum Effect (or Forer Effect)
24. Base Rate Fallacy
25. Belief Bias
26. Ben Franklin Effect
27. Berkson’s Paradox
28. Bias blind Spot
29. Bizarreness Effect

30. Boundary Extension
31. Cheerleader Effect
32. Childhood Amnesia
33. Choice-supportive Bias
34. Cognitive Dissonance
35. Commission Bias
36. Compassion Fade
37. Confirmation Bias
38. Conformity
39. Congruence Bias
40. Conjunction Fallacy (or Linda Problem)
41. Conservatism Bias (or Regressive Bias)
42. Consistency Bias
43. Context Effect
44. Continued Influence Effect
45. Contrast Effect
46. Courtesy Bias
47. Cross-race Effect
48. Cryptomnesia
49. Curse of Knowledge
50. Declinism
51. Decoy Effect
52. Default Effect
53. Defensive Attribution Hypothesis
54. Denomination Effect
55. Disposition Effect
56. Distinction Bias
57. Dread Aversion
58. Dunning-Kruger Effect
59. Duration Neglect
60. Effort Justification
61. Egocentric Bias
62. End-of-history Illusion
63. Endowment Effect
64. Escalation of Commitment (or Irrational Escalation, or Sunk Cost Fallacy)
65. Euphoric Recall
66. Exaggerated Expectation
67. Experimenter's Bias (or Expectation Bias)
68. Extension Neglect
69. Extrinsic Incentives Bias
70. Fading Affect Bias
71. Fallacy of Composition
72. Fallacy of Division
73. False Consensus Effect
74. False Memory
75. False Uniqueness Bias
76. Form Function Attribution Bias
77. Framing Effect (or Frequency Illusion, or Baader-Meinhof Phenomenon)
78. Fundamental Attribution Error
79. Gambler's Fallacy
80. Gender Bias
81. Generation Effect (or Self-generation Effect)
82. Google Effect
83. Group Attribution Error
84. Groupshift
85. Groupthink
86. Halo Effect
87. Hard-easy Effect
88. Hindsight Bias
89. Hostile Attribution Bias
90. Hot-cold Empathy Gap

91. Hot-hand Fallacy
92. Humor Effect
93. Hyperbolic Discounting
94. IKEA Effect
95. Illicit Transference
96. Illusion of Asymmetric Insight
97. Illusion of Control
98. Illusion of Explanatory Depth
99. Illusion of Transparency
100. Illusion of Validity
101. Illusory Correlation
102. Illusory Superiority
103. Illusory Truth Effect
104. Impact Bias
105. Implicit Bias
106. Information Bias
107. Ingroup Bias
108. Insensitivity To Sample Size
109. Intentionality Bias
110. Interoceptive Bias (or Hungry Judge Effect)
111. Just-world Hypothesis
112. Lag Effect
113. Less-is-better Effect
114. Leveling And Sharpening
115. Levels-of-processing Effect
116. List-length Effect
117. Logical Fallacy
118. Loss Aversion
119. Memory Inhibition
120. Mere exposure Effect (or Familiarity Principle)
121. Misattribution
122. Modality Effect
123. Money Illusion
124. Mood-congruent Memory Bias
125. Moral Credential Effect
126. Moral Luck
127. Naïve Cynicism
128. Naïve Realism
129. Negativity Bias
130. Neglect of Probability
131. Next-in-line Effect
132. Non-adaptive Choice Switching
133. Normalcy Bias
134. Not Invented Here Syndrome
135. Objectivity Illusion
136. Observer-expectancy Effect
137. Omission Bias
138. Optimism Bias
139. Ostrich Effect (or Ostrich Problem)
140. Outcome Bias
141. Outgroup Homogeneity Bias
142. Overconfidence Effect
143. Parkinson’s Law of Triviality
144. Part-list Cueing Effect
145. Peak-end Rule
146. Perky Effect
147. Pessimism Bias
148. Picture Superiority Effect
149. Placement Bias
150. Plan Continuation Bias
151. Planning Fallacy

152. Plant Blindness
153. Positivity Effect (or Socioemotional Electivity Theory)
154. Present Bias
155. Prevention Bias
156. Primacy Effect
157. Probability Matching
158. Processing Difficulty Effect
159. Pro-innovation Bias
160. Projection Bias
161. Proportionality Bias
162. Prospect Theory
163. Pseudocertainty Effect
164. Puritanical Bias
165. Pygmalion Effect
166. Reactance Theory
167. Reactive Devaluation
168. Recency Effect
169. Recency Illusion
170. Reminiscence Bump
171. Repetition Blindness
172. Restraint Bias
173. Rhyme As Reason Effect
174. Risk Compensation (or Peltzman Effect)
175. Rosy Retrospection
176. Salience Bias
177. Saying Is Believing Effect
178. Scope Neglect
179. Selection Bias
180. Self-relevance Effect
181. Self-serving Bias
182. Semmelweis Reflex
183. Serial Position Effect
184. Sexual Overperception Bias
185. Shared Information Bias
186. Social Comparison Bias
187. Social Cryptomnesia
188. Social Desirability Bias
189. Source Confusion
190. Spacing Effect
191. Spotlight Effect
192. Status Quo Bias
193. Stereotypical Bias (or Stereotype Bias)
194. Stereotyping Subadditivity Effect
195. Spacing Effect
196. Subjective Validation
197. Suffix Effect
198. Surrogation
199. Survivorship Bias
200. System Justification
201. Systematic Bias
202. Tachypsychia
203. Telescoping Effect
204. Testing Effect
205. Third-person Effect
206. Time-saving Bias
207. Tip-of-the-Tongue Phenomenon
208. Trait Ascription Bias
209. Travis Syndrome
210. Truth Bias
211. Ultimate Attribution Error
212. Unconscious Bias (or Implicit Bias)

213. Unit Bias
214. Verbatim Effect
215. Von Restorff Effect
216. Weber-Fechner Law
217. Well Traveled Road Effect
218. Women Are Wonderful Effect
219. Worse-than-average Effect
220. Zero-risk Bias
221. Zero-sum Bias

References

Acerbi, A. (2019). Cognitive attraction and online misinformation. *Palgrave Communications*, *5*(1), 15. http://dx.doi.org/10.1057/s41599-019-0224-y.

Alam, F., Shaar, S., Dalvi, F., Sajjad, H., Nikolov, A., Mubarak, H., et al. (2021). Fighting the COVID-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society. In *Findings of the association for computational linguistics: EMNLP 2021* (pp. 611–649). Punta Cana, Dominican Republic: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.findings-emnlp.56.

Alhindi, T., Petridis, S., & Muresan, S. (2018). Where is your evidence: Improving fact-checking by justification modeling. In *Proceedings of the first workshop on fact extraction and VERification (FEVER)* (pp. 85–90). Brussels, Belgium: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/W18-5513.

Armstrong, B. A., Dutescu, I. A., Tung, A., Carter, D. N., Trbovich, P. L., Wong, S., et al. (2023). Cognitive biases in surgery: Systematic review. *British Journal of Surgery*, *110*(6), 645–654. http://dx.doi.org/10.1093/bjs/znad004.

Azzopardi, L. (2021). Cognitive biases in search: A review and reflection of cognitive biases in information retrieval. In *Proceedings of the 2021 conference on human information interaction and retrieval* (pp. 27–37). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3406522.3446023.

Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM*, *61*(6), 54–61. http://dx.doi.org/10.1145/3209581.

Baeza-Yates, R. (2020). Bias in search and recommender systems. In *Proceedings of the 14th ACM conference on recommender systems* (p. 2). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3383313.3418435.

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1–24. http://dx.doi.org/10.1037/0033-2909.133.1.1.

Barnes, J. H. (1984). Cognitive biases and their impact on strategic planning. *Strategic Management Journal*, *5*(2), 129–137, URL http://www.jstor.org/stable/2486172.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*(4), 569, URL https://www.sas.upenn.edu/∼baron/papers/outcomebias.pdf.

Bhuiyan, M. M., Zhang, A. X., Sehat, C. M., & Mitra, T. (2020). Investigating differences in crowdsourced news credibility assessment: Raters, tasks, and expert criteria. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), http://dx.doi.org/10.1145/3415164.

Bollinger, B., Leslie, P., & Sorensen, A. (2011). Calorie posting in chain restaurants. *American Economic Journal: Economic Policy*, *3*(1), 91–128. http://dx.doi.org/10.1257/pol.3.1.91.

Brabazon, T. (2006). The google effect: Googling, blogging, wikis and the flattening of expertise. *International Journal of Libraries and Information Studies*, *56*(3), 157–167. http://dx.doi.org/10.1515/LIBR.2006.157.

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition*, *194*, Article 104054. http://dx.doi.org/10.1016/j.cognition.2019.104054.

Bushman, B. J. (2016). Violent media and hostile appraisals: A meta-analytic review. *Aggressive Behavior*, *42*(6), 605–613. http://dx.doi.org/10.1002/ab.21655.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. http://dx.doi.org/10.1126/science.aal4230.

Caverni, J. P., Fabre, J. M., & Gonzalez, M. (1990). *Advances in psychology, Cognitive biases*. North Holland.

Ceci, S. J., & Williams, W. M. (2020). The psychology of fact-checking. *Scientific American*, 7–13.

Ceschia, S., Roitero, K., Demartini, G., Mizzaro, S., Di Gaspero, L., & Schaerf, A. (2022). Task design in complex crowdsourcing experiments: Item assignment optimization. *Computers & Operations Research*, *148*, Article 105995. http://dx.doi.org/10.1016/j.cor.2022.105995.

Cheng, F. F., & Wu, C. S. (2010). Debiasing the framing effect: The effect of warning and involvement. *Decision Support Systems*, *49*(3), 328–334. http://dx.doi.org/10.1016/j.dss.2010.04.002.

Chou, W. Y. S., Gaysynsky, A., & Vanderpool, R. C. (2021). The COVID-19 misinfodemic: Moving beyond fact-checking. *Health Education Behavior*, *48*(1), 9–13. http://dx.doi.org/10.1177/1090198120980675.

Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., & Flammini, A. (2015). Computational fact checking from knowledge networks. *PLoS One*, *10*(6), 1–13. http://dx.doi.org/10.1371/journal.pone.0128193.

Cinelli, M., Cresci, S., Quattrociocchi, W., Tesconi, M., & Zola, P. (2022). Coordinated inauthentic behavior and information spreading on Twitter. *Decision Support Systems*, *160*, Article 113819. http://dx.doi.org/10.1016/j.dss.2022.113819.

Cinelli, M., De Francisci, G. M., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), Article e2023301118. http://dx.doi.org/10.1073/pnas.2023301118.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., et al. (2020). The COVID-19 social media infodemic. *Scientific Reports*, *10*(1), 1–10.

Clark, A. E., & Kashima, Y. (2007). Stereotypes help people connect with others in the community: A situated functional analysis of the stereotype consistency bias in communication. *Journal of Personality and Social Psychology*, *93*(6), 1028–1039. http://dx.doi.org/10.1037/0022-3514.93.6.1028.

Cooper, J. A., Gorlick, M. A., Denny, T., Worthy, D. A., Beevers, C. G., & Maddox, W. T. (2014). Training attention improves decision making in individuals with elevated self-reported depressive symptoms. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 729–741. http://dx.doi.org/10.3758/s13415-013-0220-4.

Cosmides, L., & Tooby, J. (1994). Better than rational: Evolutionary psychology and the invisible hand. *The American Economic Review*, *84*(2), 327–332, URL http://www.jstor.org/stable/2117853.

Cummings, M. (2004). Automation bias in intelligent time critical decision support systems. In *Proceedings of the AIAA 1st intelligent systems technical conference* (pp. 1–6). http://dx.doi.org/10.2514/6.2004-6313.

Das, A., Liu, H., Kovatchev, V., & Lease, M. (2023). The state of human-centered NLP technology for fact-checking. *Information Processing & Management*, *60*(2), Article 103219. http://dx.doi.org/10.1016/j.ipm.2022.103219.

Das, T., & Teng, B. S. (1999). Cognitive biases and strategic decision processes: An integrative perspective. *Journal of Management Studies*, *36*(6), 757–778. http://dx.doi.org/10.1111/1467-6486.00157.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., et al. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559. http://dx.doi.org/10.1073/pnas.1517441113.

Demartini, G., Difallah, D. E., Gadiraju, U., & Catasta, M. (2017). An introduction to hybrid human-machine information systems. *Foundations and Trends in Web Science*, *7*(1), 1–87. http://dx.doi.org/10.1561/1800000025.

Demartini, G., Mizzaro, S., & Spina, D. (2020). Human-in-the-loop artificial intelligence for fighting online misinformation: Challenges and opportunities. *IEEE Data Engineering Bulletin*, *43*(3), 65–74, URL http://sites.computer.org/debull/A20sept/p65.pdf.

Demartini, G., Roitero, K., & Mizzaro, S. (2021). Managing bias in human-annotated data: Moving beyond bias removal. *The Computing Research Repository*, arXiv:2110.13504.

Diaz, A. (2008). Through the google goggles: Sociopolitical bias in search engine design. In *Web search: multidisciplinary perspectives* (pp. 11–34). Berlin, Heidelberg: Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-540-75829-7_2.

Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and dynamics of mechanical turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining* (pp. 135–143). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3159652.3159661.

Dimara, E., Franconeri, S., Plaisant, C., Bezerianos, A., & Dragicevic, P. (2020). A task-based taxonomy of cognitive biases for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, *26*(2), 1413–1432. http://dx.doi.org/10.1109/TVCG.2018.2872577.

Draws, T., La Barbera, D., Soprano, M., Roitero, K., Ceolin, D., Checco, A., et al. (2022). The effects of crowd worker biases in fact-checking tasks. In *2022 ACM conference on fairness, accountability, and transparency* (pp. 2114–2124). Seoul, Republic of Korea: Association for Computing Machinery, http://dx.doi.org/10.1145/3531146.3534629.

Draws, T., Rieger, A., Inel, O., Gadiraju, U., & Tintarev, N. (2021). A checklist to combat cognitive biases in crowdsourcing. In *Proceedings of the ninth AAAI conference on human computation and crowdsourcing, vol. 9, no. 1* (pp. 48–59). http://dx.doi.org/10.1609/hcomp.v9i1.18939.

Drobnic Holan, A. (2018). The principles of the Truth-O-Meter: PolitiFact's methodology for independent fact-checking. https://www.politifact.com/article/2018/feb/12/principles-truth-o-meter-politifacts-methodology-i/. (Accessed: 20 June 2023).

Druică, E., Musso, F., & Ianole-Călin, R. (2020). Optimism bias during the COVID-19 pandemic: Empirical evidence from Romania and Italy. *Games*, *11*(3), http://dx.doi.org/10.3390/g11030039.

Dugan, B. (1988). Effects of assessor training on information use. *Journal of Applied Psychology*, *73*(4), 743–748. http://dx.doi.org/10.1037/0021-9010.73.4.743.

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In J. M. Olson, & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology*, *44*, 247–296. http://dx.doi.org/10.1016/B978-0-12-385522-0.00005-6.

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology*, *58*(4), 568–581. http://dx.doi.org/10.1037/0022-3514.58.4.568.

Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *SAGE Open*, *9*(1), Article 2158244019832705. http://dx.doi.org/10.1177/2158244019832705.

Eberhard, K. (2023). The effects of visualization on judgment and decision-making: A systematic literature review. *Management Review Quarterly*, *73*(1), 167–214. http://dx.doi.org/10.1007/s11301-021-00235-8.

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087–1100. http://dx.doi.org/10.3758/MC.38.8.1087.

Ehrlinger, J., Readinger, W., & Kim, B. (2016). Decision-making and cognitive biases. In H. S. Friedman (Ed.), *Encyclopedia of mental health* (2nd ed.). (pp. 5–12). Oxford: Academic Press, http://dx.doi.org/10.1016/B978-0-12-397045-9.00206-8.

Eickhoff, C. (2018). Cognitive biases in crowdsourcing. In *Proceedings of the eleventh ACM international conference on web search and data mining* (pp. 162–170). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3159652.3159654.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, *85*(5), 395–416. http://dx.doi.org/10.1037/0033-295X.85.5.395.

Epstein, Z., Pennycook, G., & Rand, D. (2020). Will the crowd game the algorithm? Using layperson judgments to combat misinformation on social media by downranking distrusted sources. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–11). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3313831.3376232.

Escola-Gascon, A., Dagnall, N., Denovan, A., Drinkwater, K., & Diez-Bosch, M. (2023). Who falls for fake news? Psychological and clinical profiling evidence of fake news consumers. *Personality and Individual Differences*, *200*, http://dx.doi.org/10.1016/j.paid.2022.111893.

Etchells, P. (2015). Declinism: Is the world actually getting worse. *The Guardian*, *15*, 1087–1089.

Evans, N., Edge, D., Larson, J., & White, C. (2020). News provenance: Revealing news text reuse at web-scale in an augmented news search experience. In *CHI EA '20, Extended abstracts of the 2020 CHI conference on human factors in computing systems* (pp. 1–8). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3334480.3375225.

FactCheck. org (2020). Our process. https://www.factcheck.org/our-process/. (Accessed: 15 December 2021).

Ferreira, W., & Vlachos, A. (2016). Emergent: a novel data-set for stance classification. In K. Knight, A. Nenkova, & O. Rambow (Eds.), *NAACL HLT 2016, the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1163–1168). The Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/n16-1138.

Fichten, C. S., & Sunerton, B. (1983). Popular horoscopes and the "Barnum Effect". *Journal of Psychology*, *114*(1), 123–134. http://dx.doi.org/10.1080/00223980.1983.9915405.

Fisher, K. L., & Statman, M. (2000). Cognitive biases in market forecasts. *The Journal of Portfolio Management*, *27*(1), 72–81. http://dx.doi.org/10.3905/jpm.2000.319785.

Follett, K. J., & Hess, T. M. (2002). Aging, cognitive complexity, and the fundamental attribution error. *The Journals of Gerontology: Series B*, *57*(4), P312–P323. http://dx.doi.org/10.1093/geronb/57.4.P312.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42. http://dx.doi.org/10.1016/j.socec.2010.10.008.

Gadiraju, U., Yang, J., & Bozzon, A. (2017). Clarity is a worthwhile quality: On the role of task clarity in microtask crowdsourcing. In *Proceedings of the 28th ACM conference on hypertext and social media* (pp. 5–14). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3078714.3078715.

Gigerenzer, G., & Selten, R. (2008). Bounded and rational. In *Philosophie: grundlagen und anwendungen/philosophy: foundations and applications* (pp. 233–257). Brill | mentis, http://dx.doi.org/10.30965/9783969750056_016.

Gillier, T., Chaffois, C., Belkhouja, M., Roth, Y., & Bayus, B. L. (2018). The effects of task instructions in crowdsourcing innovative ideas. *Technological Forecasting and Social Change*, *134*, 35–44. http://dx.doi.org/10.1016/j.techfore.2018.05.005.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2011). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, *19*(1), 121–127. http://dx.doi.org/10.1136/amiajnl-2011-000089.

Gomroki, G., Behzadi, H., Fattahi, R., & Fadardi, J. S. (2023). Identifying effective cognitive biases in information retrieval. *J. Inf. Sci.*, *49*(2), 348–358. http://dx.doi.org/10.1177/01655515211001777.

Groome, D., & Eysenck, M. (2016). *An introduction to applied cognitive psychology*. Psychology Press.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12*(4), 392–407. http://dx.doi.org/10.1016/S0022-1031(76)80006-6.

Harvey, J. H., Town, J. P., & Yarkin, K. L. (1981). How fundamental is "The Fundamental Attribution Error"? *Journal of Personality and Social Psychology*, *40*(2), 346–349. http://dx.doi.org/10.1037/0022-3514.40.2.346.

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, *10*(1), 47–66. http://dx.doi.org/10.1207/s15327957pspr1001_3.

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The evolution of cognitive bias. In *The handbook of evolutionary psychology* (pp. 1–20). John Wiley & Sons, Ltd, http://dx.doi.org/10.1002/9781119125563.evpsych241.

Hassan, N., Adair, B., Hamilton, J. T., Li, C., Tremayne, M., Yang, J., et al. (2015). The quest to automate fact-checking. In *Proceedings of the 2015 computation + journalism symposium* (pp. 1–5). URL http://cj2015.brown.columbia.edu/papers/automate-fact-checking.pdf.

Hayibor, S., & Wasieleski, D. M. (2009). Effects of the use of the availability heuristic on ethical decision-making in organizations. *Journal of Business Ethics*, *84*, 151–165, URL http://www.jstor.org/stable/40294779.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, *32*, 113–135. http://dx.doi.org/10.1016/j.riob.2012.11.003.

Hettiachchi, D., Kostakos, V., & Goncalves, J. (2022). A survey on task assignment in crowdsourcing. *ACM Computing Surveys*, *55*(3), http://dx.doi.org/10.1145/3494522.

Hettiachchi, D., Schaekermann, M., McKinney, T. J., & Lease, M. (2021). The challenge of variable effort crowdsourcing and how visible gold can help. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), http://dx.doi.org/10.1145/3476073.

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychology Bullettin*, *138*(2), 211–237. http://dx.doi.org/10.1037/a0025940.

Hom, H. L. (2022). Perspective-taking and hindsight bias: When the target is oneself and/or a peer. *Current Psychology*, http://dx.doi.org/10.1007/s12144-021-02413-z.

Javdani, M., & Chang, H. J. (2023). Who said or what said? Estimating ideological bias in views among economists. *Cambridge Journal of Economics*, *47*(2), 309–339. http://dx.doi.org/10.1093/cje/beac071.

Jerit, J. (2008). Issue framing and engagement: Rhetorical strategy in public policy debates. *Political Behavior*, *30*(1), 1–24, URL http://www.jstor.org/stable/40213302.

Johnson, D. D., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution*, *28*(8), 474–481. http://dx.doi.org/10.1016/j.tree.2013.05.014.

Jones, E. L. (1963). The courtesy bias in south-east Asian surveys. In *Social research in developing countries: surveys and censuses in the third world International Social Science Journal*, 70–76, URL https://unesdoc.unesco.org/ark:/48223/pf0000016829.

Kafaee, M., Marhamati, H., & Gharibzadeh, S. (2021). "Choice-supportive Bias" in science: Explanation and mitigation. *Accountability in Research*, *28*(8), 528–543. http://dx.doi.org/10.1080/08989621.2021.1872377.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Macmillan.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and biases: the psychology of intuitive judgment* (pp. 49–81). Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511808098.004.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251. http://dx.doi.org/10.1037/h0034747.

Karduni, A., Wesslen, R., Santhanam, S., Cho, I., Volkova, S., Arendt, D., et al. (2018). Can you verifi this? Studying uncertainty and decision-making about misinformation using visual analytics. In *Proceedings of the international AAAI conference on web and social media*: *vol. 12*, (pp. 1–10). http://dx.doi.org/10.1609/icwsm.v12i1.15014.

Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, *38*(2), 95–115. http://dx.doi.org/10.1007/s11166-009-9060-6.

Kazdin, A. E. (1977). Artifact, bias, and complexity of assessment: The ABCs of reliability. *Journal of Applied Behavior Analysis*, *10*(1), 141–150. http://dx.doi.org/10.1901/jaba.19P77.10-141.

Kiesel, J., Spina, D., Wachsmuth, H., & Stein, B. (2021). The meant, the said, and the understood: Conversational argument search and cognitive biases. In *Proceedings of the 3rd conference on conversational user interfaces* (pp. 1–5). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3469595.3469615.

Kim, S., Goldstein, D., Hasher, L., & Zacks, R. T. (2005). Framing effects in Younger and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(4), P215–P218. http://dx.doi.org/10.1093/geronb/60.4.p215.

Kiss, Á., & Simonovits, G. (2014). Identifying the bandwagon effect in two-round elections. *Public Choice*, *160*(3), 327–344. http://dx.doi.org/10.1007/s11127-013-0146-y.

Kupfer, C., Prassl, R., Fleiß, J., Malin, C., Thalmann, S., & Kubicek, B. (2023). Check the box! How to deal with automation bias in AI-based personnel selection. *Frontiers in Psychology*, *14*, http://dx.doi.org/10.3389/fpsyg.2023.1118723.

Kuran, T., & Sunstein, C. R. (1998). Availability cascades and risk regulation. *Stanford Law Review*, *51*, URL https://ssrn.com/abstract=138144.

La Barbera, D., Roitero, K., Demartini, G., Mizzaro, S., & Spina, D. (2020). Crowdsourcing truthfulness: The impact of judgment scale and assessor bias. In J. M. Jose, E. Yilmaz, J. a. Magalhães, P. Castells, N. Ferro, M. J. Silva, & et al. (Eds.), *Advances in information retrieval* (pp. 207–214). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-45442-5_26.

Lee, S. H., & Lee, K. T. (2023). The impact of pandemic-related stress on attentional bias and anxiety in alexithymia during the COVID-19 pandemic. *Scientific Reports*, *13*(1), 6327. http://dx.doi.org/10.1038/s41598-023-33326-5.

Leighton, J. P., & Sternberg, R. J. (2004). *The nature of reasoning*. UK: Cambridge University Press.

Leman, P. J., & Cinnirella, M. (2007). A major event has a major cause: Evidence for the role of heuristics in reasoning about conspiracy theories. *Social Psychological Review*, *9*(2), 18–28. http://dx.doi.org/10.53841/bpsspr.2007.9.2.18.

Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin*, *85*(5), 1030–1051. http://dx.doi.org/10.1037/0033-2909.85.5.1030.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. http://dx.doi.org/10.1177/1529100612451018.

Li, G., Dong, M., Yang, F., Zeng, J., Yuan, J., Jin, C., et al. (2020). Misinformation-oriented expert finding in social networks. *World Wide Web*, *23*(2), 693–714. http://dx.doi.org/10.1007/s11280-019-00717-6.

Lievens, F. (2001). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology*, *86*(2), 255. http://dx.doi.org/10.1037/0021-9010.86.2.255.

Lind, M., Visentini, M., Mäntylä, T., & Del Missier, F. (2017). Choice-supportive misremembering: A new taxonomy and review. *Frontiers in Psychology*, *8*, http://dx.doi.org/10.3389/fpsyg.2017.02062.

Lindgren, E., Lindholm, T., Vliegenthart, R., Boomgaarden, H. G., Damstra, A., Strömbäck, J., et al. (2022). Trusting the facts: The role of framing, news media as a (trusted) source, and opinion resonance for perceived truth in statistical statements. *Journalism & Mass Communication Quarterly*, Article 10776990221117117. http://dx.doi.org/10.1177/10776990221117117.

Liu, Y., & Wu, Y. F. B. (2020). FNED: A deep network for fake news early detection on social media. *ACM Transactions on Information Systems*, *38*(3), http://dx.doi.org/10.1145/3386253.

Luo, G. Y. (2014). Conservatism bias and asset price overreaction or underreaction to new information in a competitive securities market. In *Asset price response to new information: the effects of conservatism bias and representativeness heuristic* (pp. 5–14). New York, NY: Springer New York, http://dx.doi.org/10.1007/978-1-4614-9369-3_2.

Lurie, E., & Mustafaraj, E. (2018). Investigating the effects of google's search engine result page in evaluating the credibility of online news sources. In *Proceedings of the 10th ACM conference on web science* (pp. 107–116). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3201064.3201095.

MacCoun, R. J. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology*, *49*(1), 259–287. http://dx.doi.org/10.1146/annurev.psych.49.1.259.

Malenka, D. J., Baron, J. A., Johansen, S., Wahrenberger, J. W., & Ross, J. M. (1993). The framing effect of relative and absolute risk. *Journal of General Internal Medicine*, *8*(10), 543–548. http://dx.doi.org/10.1007/BF02599636.

Mastroianni, A. M., & Gilbert, D. T. (2023). The illusion of moral decline. *Nature*, http://dx.doi.org/10.1038/s41586-023-06137-x.

Matute, H., Yarritu, I., & Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology*, *102*(3), 392–405. http://dx.doi.org/10.1348/000712610X532210.

Mena, P. (2019). Principles and boundaries of fact-checking: Journalists' perceptions. *Journalism Practice*, *13*(6), 657–672. http://dx.doi.org/10.1080/17512786.2018.1547655.

Moher, D., Cook, D. J., Eastwood, S., Olkinm, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *The Lancet*, *354*(9193), 1896–1900. http://dx.doi.org/10.1016/S0140-6736(99)04149-5.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, *339*, http://dx.doi.org/10.1136/bmj.b2535.

Mowshowitz, A., & Kawaguchi, A. (2005). Measuring search engine bias. *Information Processing & Management*, *41*(5), 1193–1205. http://dx.doi.org/10.1016/j.ipm.2004.05.005.

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, *22*(2), 103–122. http://dx.doi.org/10.1002/ejsp.2420220202.

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, *26*(9), 1142–1150. http://dx.doi.org/10.1177/01461672002611010.

Nesse, R. M. (2001). Natural selection and the regulation of defensive responses. *Annals of the New York Academy of Sciences*, *935*, 75–85, URL https://pubmed.ncbi.nlm.nih.gov/11411177/.

Nesse, R. M. (2005). Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector principle. *Evolution and Human Behaviour*, *26*(1), 88–105. http://dx.doi.org/10.1016/j.evolhumbehav.2004.08.002.

Newman, M. C., Schwarz, N., & Ly, D. P. (2020). Truthiness, the illusory truth effect, and the role of need for cognition. *Consciousness and Cognition*, *78*, Article 102866. http://dx.doi.org/10.1016/j.concog.2019.102866.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, *17*(2), 141–161. http://dx.doi.org/10.1017/epi.2018.32.

Ni, F., Arnott, D., & Gao, S. (2019). The anchoring effect in business intelligence supported decision-making. *Journal of Decision Systems*, *28*(2), 67–81. http://dx.doi.org/10.1080/12460125.2019.1620573.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. http://dx.doi.org/10.1037/1089-2680.2.2.175.

Oeberst, A., & Imhoff, R. (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), 1464–1487. http://dx.doi.org/10.1177/17456916221148147.

Oeldorf-Hirsch, A., & DeVoss, C. L. (2020). Who posted that story? Processing layered sources in facebook news posts. *Journalism & Mass Communication Quarterly*, *97*(1), 141–160. http://dx.doi.org/10.1177/1077699019857673.

Otterbacher, J., Checco, A., Demartini, G., & Clough, P. (2018). Investigating user perception of gender bias in image search: The role of sexism. In *The 41st international ACM SIGIR conference on research & development in information retrieval* (pp. 933–936). Association for Computing Machinery, http://dx.doi.org/10.1145/3209978.3210094.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, http://dx.doi.org/10.1136/bmj.n71.

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*, http://dx.doi.org/10.1136/bmj.n160.

Park, S., Park, J. Y., & Kang, J. h. (2021). The presence of unexpected biases in online fact-checking. *Harvard Kennedy School Misinformation Review*, *2*, http://dx.doi.org/10.37016/mr-2020-53.

Pell, G., Homer, M. S., & Roberts, T. E. (2008). Assessor training: Its effects on criterion-based assessment in a medical context. *International Journal of Research & Method in Education*, *31*(2), 143–154. http://dx.doi.org/10.1080/17437270802124525.

Pitts, J., Coles, C., Thomas, P., & Smith, F. (2002). Enhancing reliability in portfolio assessment: Discussions between assessors. *Medical Teacher*, *24*(2), 197–201. http://dx.doi.org/10.1080/01421590220125321.

Pornari, C. D., & Wood, J. (2010). Peer and cyber aggression in secondary school students: The role of moral disengagement, hostile attribution bias, and outcome expectancies. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *36*(2), 81–94. http://dx.doi.org/10.1002/ab.20336.

Porter, J. (2020). Snopes forced to scale back fact-checking. https://www.theverge.com/2020/3/24/21192206/snopes-coronavirus-covid-19-misinformation-fact-checking-staff. (Accessed: 20 June 2023).

Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proceedings of the National Academy of Sciences*, *118*(37), Article e2104235118. http://dx.doi.org/10.1073/pnas.2104235118.

Ralston, R. (2022). Make us great again: The causes of declinism in major powers. *Security Studies*, *31*(4), 667–702. http://dx.doi.org/10.1080/09636412.2022.2133626.

Reimer, T., Reimer, A., & Czienskowski, U. (2010). Decision-making groups attenuate the discussion bias in favor of shared information: A meta-analysis. *Communication Monographs*, *77*(1), 121–142. http://dx.doi.org/10.1080/03637750903514318.

Reis, J. C. S., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Explainable machine learning for fake news detection. In *Proceedings of the 10th ACM conference on web science* (pp. 17–26). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3292522.3326027.

Ries, A. (2006). Understanding marketing psychology and the halo effect. *Advertising Age*, *17*, URL https://adage.com/article/al-ries/understanding-marketing-psychology-halo-effect/108676.

Robson, D. (2019). The bias that can cause catastrophe. https://www.bbc.com/worklife/article/20191001-the-bias-behind-the-worlds-greatest-catastrophes/. (Accessed: 06 July 2023).

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426. http://dx.doi.org/10.1177/1745691612454303.

Roitero, K., Demartini, G., Mizzaro, S., & Spina, D. (2018). How many truth levels? Six? One hundred? Even more? Validating truthfulness of statements via crowdsourcing. In *Proceedings of the CIKM 2018 workshops co-located with 27th ACM international conference on information and knowledge management* (pp. 1–6). URL http://ceur-ws.org/Vol-2482/paper38.pdf.

Roitero, K., Soprano, M., Fan, S., Spina, D., Mizzaro, S., & Demartini, G. (2020). Can the crowd identify misinformation objectively? The effects of judgment scale and assessor's background. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval* (pp. 439—448). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3397271.3401112.

Roitero, K., Soprano, M., Portelli, B., Spina, D., Della Mea, V., Serra, G., et al. (2020). The COVID-19 infodemic: Can the crowd judge recent misinformation objectively? In *Proceedings of the 29th ACM international conference on information & knowledge management* (pp. 1305–1314). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3340531.3412048.

Rubin, Z., & Peplau, L. A. (1975). Who believes in a just world? *Journal of Social Issues*, *31*(3), 65–89. http://dx.doi.org/10.1111/j.1540-4560.1975.tb00997.x.

Ruffo, G., Semeraro, A., Giachanou, A., & Rosso, P. (2023). Studying fake news spreading, polarisation dynamics, and manipulation by bots: A tale of networks and language. *Computer Science Review*, *47*, Article 100531. http://dx.doi.org/10.1016/j.cosrev.2022.100531.

Schul, Y. (1993). When warning succeeds: The effect of warning on success in ignoring invalid information. *Journal of Experimental Social Psychology*, *29*(1), 42–62. http://dx.doi.org/10.1006/jesp.1993.1003.

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), R941–R945. http://dx.doi.org/10.1016/j.cub.2011.10.030.

Shatz, I. (2020a). The availability cascade: How information spreads on a large scale. https://effectiviology.com/availability-cascade/. (Accessed: 20 June 2023).

Shatz, I. (2020b). The bandwagon effect: Why people tend to follow the crowd. https://effectiviology.com/bandwagon/. (Accessed: 15 December 2021).

Shin, J., & Thorson, K. (2017). Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *Journal of Communication*, *67*(2), 233–255. http://dx.doi.org/10.1111/jcom.12284.

Simonite, T. (2018). When it comes to Gorillas, google photos remains blind. https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/. (Accessed: 20 June 2023).

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177*(3), 1333–1352. http://dx.doi.org/10.1016/j.ejor.2005.04.006.

Soprano, M., Roitero, K., La Barbera, D., Ceolin, D., Spina, D., Mizzaro, S., et al. (2021). The many dimensions of truthfulness: Crowdsourcing misinformation assessments on a multidimensional scale. *Information Processing & Management*, *58*(6), Article 102710. http://dx.doi.org/10.1016/j.ipm.2021.102710.

Spina, D., Sanderson, M., Angus, D., Demartini, G., McKay, D., Saling, L. L., et al. (2023). Human-AI cooperation to tackle misinformation and polarization. *Communications of the ACM*, *66*(7), 40–45. http://dx.doi.org/10.1145/3588431.

Stall, L. M., & Petrocelli, J. V. (2023). Countering conspiracy theory beliefs: Understanding the conjunction fallacy and considering disconfirming evidence. *Applied Cognitive Psychology*, *37*(2), 266–276. http://dx.doi.org/10.1002/acp.3998.

Stubenvoll, M., & Matthes, J. (2022). Why retractions of numerical misinformation fail: The anchoring effect of inaccurate numbers in the news. *Journalism & Mass Communication Quarterly*, *99*(2), 368–389. http://dx.doi.org/10.1177/10776990211021800.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*(1), 1–26. http://dx.doi.org/10.1111/1529-1006.001.

Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition*, *9*(3), 286–299. http://dx.doi.org/10.1016/j.jarmac.2020.06.006.

Sylvia Chou, W. Y., Gaysynsky, A., & Cappella, J. N. (2020). Where we go from here: Health misinformation on social media. *American Journal of Public Health*, *110*(S3), S273–S275. http://dx.doi.org/10.2105/AJPH.2020.305905.

Szpara, M. Y., & Wylie, C. E. (2005). National board for professional teaching standards assessor training: Impact of bias reduction exercises. *Teachers College Record*, *107*(4), 803–841. http://dx.doi.org/10.1177/016146810510700410.

The RMIT ABC Fact Check Team (2021). Fact check. https://www.abc.net.au/news/factcheck/about/. (Accessed: 20 June 2023).

Thomas, O. (2018). Two decades of cognitive bias research in entrepreneurship: What do we know and where do we go from here? *Management Review Quarterly*, *68*(2), 107–143. http://dx.doi.org/10.1007/s11301-018-0135-9.

Thomas, E. F., Cary, N., Smith, L. G., Spears, R., & McGarty, C. (2018). The role of social media in shaping solidarity and compassion fade: How the death of a child turned apathy into action but distress took it away. *New Media & Society*, *20*(10), 3778–3798. http://dx.doi.org/10.1177/1461444818760819.

Thompson, C. P., Skowronski, J. J., & Lee, D. J. (1988). Telescoping in dating naturally occurring events. *Memory & Cognition*, *16*(5), 461–468. http://dx.doi.org/10.3758/BF03214227.

Thorne, J., & Vlachos, A. (2018). Automated fact checking: Task formulations, methods and future directions. In *Proceedings of the 27th international conference on computational linguistics* (pp. 3346–3359). Santa Fe, New Mexico, USA: Association for Computational Linguistics, URL https://aclanthology.org/C18-1283.

Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, *23*(5), 727–741. http://dx.doi.org/10.1017/S0140525X00003447.

Traberg, C. S., & Van Der Linden, S. (2022). Birds of a feather are persuaded together: Perceived source credibility mediates the effect of political bias on misinformation susceptibility. *Personality and Individual Differences*, *185*, Article 111269. http://dx.doi.org/10.1016/j.paid.2021.111269.

Tucker, J. A., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., et al. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Social Science Research Network*, http://dx.doi.org/10.2139/ssrn.3144139, URL https://ssrn.com/abstract=3144139.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293. http://dx.doi.org/10.1037/0033-295X.90.4.293.

Vlachos, A., & Riedel, S. (2014). Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 workshop on language technologies and computational social science* (pp. 18–22). Baltimore, MD, USA: Association for Computational Linguistics, http://dx.doi.org/10.3115/v1/W14-2508.

Vyas, K., Murphy, D., & Greenberg, N. (2023). Cognitive biases in military personnel with and without PTSD: a systematic review. *Journal of Mental Health*, *32*(1), 248–259. http://dx.doi.org/10.1080/09638237.2020.1766000, PMID: 32437214.

Wabnegger, A., Gremsl, A., & Schienle, A. (2021). The association between the belief in coronavirus conspiracy theories, miracles, and the susceptibility to conjunction fallacy. *Applied Cognitive Psychology*, *35*(5), 1344–1348. http://dx.doi.org/10.1002/acp.3860.

Walter, N., Cohen, J., Lance Holbert, R., & Morag, J. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(4), 350–375. http://dx.doi.org/10.1080/10584609.2019.1668894.

Wang, W. Y. (2017). "Liar, Liar Pants on Fire": A new benchmark dataset for fake news detection. In R. Barzilay, & M. Kan (Eds.), *Proceedings of the 55th annual meeting of the association for computational linguistics*: *vol. 4*, (pp. 422–426). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/P17-2067.

Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine*, *240*, Article 112552. http://dx.doi.org/10.1016/j.socscimed.2019.112552.

Weiss, D., & Taskar, B. (2010). Structured prediction cascades. In Y. W. Teh, & M. Titterington (Eds.), *Proceedings of machine learning research*: *vol. 9*, *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 916–923). Chia Laguna Resort, Sardinia, Italy: PMLR, URL https://proceedings.mlr.press/v9/weiss10a.html.

Welsh, M. B., & Navarro, D. J. (2012). Seeing is believing: Priors, trust, and base rate neglect. *Organizational Behavior and Human Decision Processes*, *119*(1), 1–14. http://dx.doi.org/10.1016/j.obhdp.2012.04.001.

Wesslen, R., Santhanam, S., Karduni, A., Cho, I., Shaikh, S., & Dou, W. (2019). Investigating effects of visual anchors on decision-making about misinformation. *Computer Graphics Forum*, *38*(3), 161–171. http://dx.doi.org/10.1111/cgf.13679.

Wilkie, C., & Azzopardi, L. (2014). Best and fairest: An empirical analysis of retrieval system bias. In M. de Rijke, T. Kenter, A. P. de Vries, C. Zhai, F. de Jong, K. Radinsky, & et al. (Eds.), *Advances in information retrieval* (pp. 13–25). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-06028-6_2.

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *41*(1), 135–163. http://dx.doi.org/10.1007/s11109-018-9443-y.

Wood, J. R., & Wood, L. E. (2008). Card sorting: Current practices and beyond. *Journal of Usability Studies*, *4*(1), 1–6. http://dx.doi.org/10.5555/2835577.2835578.

Wu, L., Rao, Y., Yang, X., Wang, W., & Nazir, A. (2020). Evidence-aware hierarchical interactive attention networks for explainable claim verification. In C. Bessiere (Ed.), *Proceedings of the twenty-ninth international joint conference on artificial intelligence* (pp. 1388–1394). International Joint Conferences on Artificial Intelligence Organization, http://dx.doi.org/10.24963/ijcai.2020/193, Main track.

Zheng, L., Cui, P., Li, X., & Huang, R. (2018). Synchronous discussion between assessors and assessees in web-based peer assessment: Impact on writing performance, feedback quality, meta-cognitive awareness and self-efficacy. *Assessment & Evaluation in Higher Education*, *43*(3), 500–514. http://dx.doi.org/10.1080/02602938.2017.1370533.

Zhou, Y., & Shen, L. (2022). Confirmation bias and the persistence of misinformation on climate change. *Communication Research*, *49*(4), 500–523.

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, *53*(5), http://dx.doi.org/10.1145/3395046.

Zollo, F. (2019). Dealing with digital misinformation: A polarised context of narratives and tribes. *EFSA Journal*, *17*(S1), Article e170720. http://dx.doi.org/10.2903/j.efsa.2019.e170720.

Zollo, F., & Quattrociocchi, W. (2018). Misinformation spreading on facebook. In *Complex spreading phenomena in social systems: influence and contagion in real-world social networks* (pp. 177–196). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-77332-2_10.