

## Does fact-checking work?

### A qualitative critical review of evidence, limits, and conditions

  **João Pedro Baptista**

University of Trás-os-Montes and  
Alto Douro  
Vila Real, Portugal

 **Francisco Conrado Filho**

University of Minho  
Braga, Portugal

 **Anabela Gradim**

University of Beira Interior  
Covilhã, Portugal

#### Abstract

This article offers a qualitative critical review of the research on the effectiveness of fact-checking as a response to disinformation, covering the period between 2010 and 2025. Although journalism constitutes the historical core and the normative reference of the practice, the review deliberately treats fact-checking as a sociotechnical intervention whose boundaries exceed the newsroom, encompassing correction formats, preventive inoculation, crowdsourced verification, and artificial-intelligence support. Drawing on 148 studies identified through citation tracking and organised by qualitative synthesis, the analysis delineates three phases of scholarly understanding: an initial phase of skepticism about corrective effects; a phase of reassessment that establishes a conditional but real effectiveness and relativises the backfire effect; and a recent phase centred on explanatory formats, prevention, and automation. Across these phases, the evidence converges on the view that fact-checking reduces misperceptions, though its impact remains conditioned by motivated political reasoning, identity attachment, and uneven levels of media and information literacy. The article argues that fact-checking should be understood as a multidimensional practice that articulates cognitive, social, and technological strategies, and it draws out the implications for newsroom routines, fact-checking organisations, and professional training, with particular emphasis on combining design-based corrections with community engagement. By connecting theoretical debate with professional challenges, the review mobilises fact-checking as an entry point into wider questions of journalistic ethics and digital platform governance.

Keywords: fact-checking, journalism practices, disinformation, news verification, media literacy, audience engagement, inoculation theory, AI-assisted verification

#### 1. Introduction

A critical understanding of the discursive dynamics of society has always been one of the main challenges in the field of communication. With the emergence of the phenomenon of disinformation, now amplified by social algorithms, it has become a focal point of analysis. First and foremost, because disinformation cannot be understood in isolation and must instead be inscribed within a broader conceptual framework described as post-truth. This concept highlights not only an indifference towards objective truth, but also a sociocultural and communicational context. In this context, factual objectivity is systematically subordinated to personal beliefs, political identities, and sophisticated rhetorical strategies and affects (McIntyre, 2018; Lewandowsky et al., 2017; Harsin, 2015). Facts lose a significant part of their symbolic and functional value, being instrumentalized as mere rhetorical elements in political and identity disputes.

It is precisely within this field, densely intertwined with symbolic disputes, that fact-checking asserts itself as a practice. Emerging primarily in the first decade of the 21st century, fact-checking crystallizes a dual claim. It is an affiliation with the classic journalistic ethos, anchored in systematic and transparent verification procedures, but it is also a vocation as a sociotechnical movement, positioning itself as a normative response to the erosion of informative integrity and discursive responsibility in contemporary public space (Graves,

2016; Amazeen, 2020; Cheruiyot & Ferrer-Conill, 2018). This practice, far from being univocal or free from ambivalence, is permeated by constitutive tensions between the aspiration for objectivity and the constraints of a public sphere saturated with polarization, between the commitment to factual truth and the inevitability of its politicized reading (Uscinski & Butler, 2013; Nieminen & Rapeli, 2018).

This work questions the effectiveness of fact-checking as a discursive device. In a research field marked by disparate and sometimes contradictory results, it is crucial to identify the factors that condition its effectiveness and understand the conditions for this practice to produce tangible effects in reversing false beliefs. Multiple experimental studies indicate that, in contexts with less ideological charge or under conditions of attenuated polarization control, exposure to fact-checked content can foster cognitive reevaluations and orient recipients toward positions more in line with empirical evidence (Nyhan & Reifler, 2010; Walter & Murphy, 2018). But beyond these, the literature has also documented paradoxical scenarios in which fact-checking not only fails to correct, but reinforces pre-existing beliefs, especially when these are intertwined with identity affiliations or deeply rooted moral convictions, a phenomenon described as the backfire effect (Lewandowsky et al., 2012; Swire-Thompson et al., 2023).

The circulation and consumption of fact-checked content is embedded in densely stratified digital dynamics, permeated by logics of political polarization and ideological segmentation that structure how individuals access, evaluate, and react to factual information. Far from being evenly distributed in the public space, fact-checks circulate according to selective patterns, often confined to communities already predisposed to their positive reception (Lyons et al. 2020). Recent research has revealed that audiences who regularly consume fact-checked content generally share a set of epistemic and cultural predispositions, including greater media literacy, analytical disposition, and a normative commitment to factuality, which distinguishes them from segments more vulnerable to disinformation (Pennycook & Rand, 2019; Swire-Thompson & Lazer, 2020).

This consumption profile, while encouraging from the perspective of informational responsibility, also highlights the structural limits of fact-checking as a strategy for mitigating epistemic disorder. Indeed, the concentration of fact-check consumption in cognitive niches already oriented towards the valorization of verifiable truth reveals the difficulty in penetrating discursive zones where distrust of journalistic mediation, skepticism toward institutions, and identity affinity with false narratives constitute nearly insurmountable barriers. Instead of expanding consensus, fact-checking runs the risk of reinforcing cleavages, being appropriated as a validation tool by those who already share its assumptions and summarily rejected by audiences whose worldview is dissonant (Pennycook & Rand, 2019).

This asymmetrical consumption is largely explained by cognitive and motivational mechanisms that regulate how individuals process information, especially in contexts with a high ideological load. Confirmation bias, the tendency to seek out, interpret, and share content that revalidates already established beliefs, acts as a selective filter, amplifying the reception of fact-checks consistent with the recipient's values and discarding, if not hostile, those that challenge their political or identity orientation (Lewandowsky, Ecker & Cook, 2017). In this sense, fact-checks do not operate in a rational vacuum, but are read and appropriated within complex affective-ideological landscapes, where the acceptance of factual truth depends not only on the quality of the evidence presented, but also on its symbolic compatibility with the recipient's mental and affective frameworks (Lyons et al., 2020).

In this context, ideological polarization plays a crucial role in the circulation of fact-checked content. Digital platforms often amplify this polarization through algorithms that prioritize emotionally charged content or content that generates greater engagement (Westlund, Bélair-Gagnon, Graves, Larsen & Steensen, 2024). This dynamic contributes to fact-checking circulating predominantly within ideologically homogeneous groups, reducing its effectiveness in reaching divergent audiences. In short, in highly polarized environments, fact-checks face increased difficulties in penetrating opposing "information bubbles," often further reinforcing media and social fragmentation (Swire-Thompson & Lazer, 2020).

This dynamic ultimately challenges the practice and effectiveness of fact-checking as a sociotechnical intervention against disinformation. The critical integration of these dimensions (cognitive, motivational, social, and technological) requires refined and targeted strategies capable of overcoming the current limitations in the consumption and circulation of information. Before such strategies can be developed, however, it is essential to understand these limitations and their present effects. To this end, we aim to critically examine the question of fact-checking's effectiveness through a primarily theoretical review that seeks to articulate different perspectives and empirical results.

In addition to addressing the effectiveness of fact-checking, this article essentially seeks to contribute to the theoretical development of the field, fostering an academic discussion around the topic that goes beyond the binary question imposed on fact-checking, namely, whether or not it has a corrective effect. The truth is that several recent empirical findings suggest that effectiveness should no longer be understood as an intrinsic property of fact-checking itself, but rather should be understood or analyzed as the result of consequential interactions between various cognitive, political, technological, and contextual factors (Lewandowsky et al., 2017; Ecker et al., 2020; Walter et al., 2020; Swire-Thompson & Lazer, 2020). Therefore, the originality of this research lies not only in the collection of evidence published between 2010 and 2025, but also in the integration of scattered findings into a common interpretative framework, capable of explaining why studies often reach different conclusions when examining the same phenomenon.

A clarification of scope is warranted before proceeding, since the object of this review is deliberately broader than journalistic verification in the strict sense. Fact-checking emerged from, and remains anchored in, the professional ethos of journalism, and journalistic organisations continue to provide its institutional core and its most visible public face. Yet the practice has progressively exceeded the newsroom, articulating itself with cognitive, computational, and participatory interventions that share its corrective purpose without necessarily sharing its professional authorship. Crowdsourced verification, automated and AI-assisted labelling, and inoculation-based prevention are not journalism in the conventional sense, but they form part of the same sociotechnical effort to counter disinformation and are routinely discussed in dialogue with journalistic fact-checking. We therefore adopt an inclusive understanding of fact-checking as a multidimensional and sociotechnical practice of which journalism is the centre of gravity rather than the entire territory. This choice is not merely terminological: it determines the composition of the corpus and explains why studies whose object is preventive or non-professional correction are read here alongside studies of newsroom-based verification, as complementary expressions of a common problem rather than as a single homogeneous genre.

Against this background, the review is structured by three core questions:

Q1: To what extent is fact-checking effective in reducing false beliefs and combating disinformation?

Q2: What specific factors – whether cognitive, political, technological, or contextual – determine and condition the effectiveness or ineffectiveness of this type of discursive intervention?

Q3: What conceptual trends emerge from the literature regarding the evolution of fact-checking effectiveness research between 2010 and 2025?

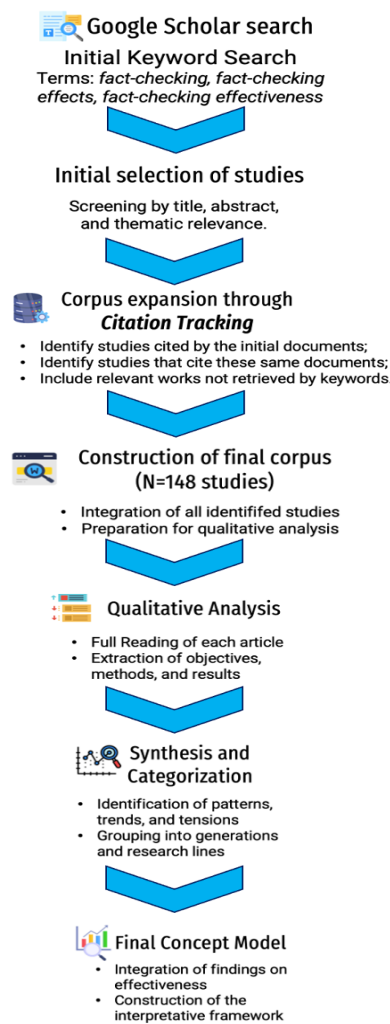
Whereas the first two questions address the phenomenon of fact-checking itself, the third operates at a different analytical level, asking how the scholarly field has framed and reframed that phenomenon over time. By exploring these questions, this study clarifies the role of fact-checking in contemporary societies and offers not only a critical mapping of its guiding principles, but also identifies important literature gaps.

## 2. Methodology

The literature review was conducted based on a mapping of scientific publications that included both direct research results and citations linked to each other (citation tracking). First and foremost, the choice of Google Scholar (GS) as the main starting point for this collection was justified by its comprehensiveness. GS covers not only journals indexed in traditional databases but also gray literature, working papers, and preprints, sometimes at different stages of publication. These sources expanded the scope of the review by incorporating emerging evidence and early-stage impact studies. This approach is not new, and other authors have pointed out that in fields such as communication and social sciences, this breadth mitigates reliance on databases restricted by editorial criteria and paywalls (Haddaway et al., 2015; Gusenbauer, 2019). This approach also made it possible to account for research circulating in regional contexts or in open access publications that might otherwise be underrepresented in the review (Delgado López-Cózar & Repiso, 2013).

The initial search used a combination of descriptors related to the effectiveness of fact-checking, its effects, and the impact of corrections on public opinion ("fact checking, fact-checking evidence, fact-checking effects")(Figure 1). Subsequently, the corpus was expanded through citation chaining, identifying studies that referenced previously located work. This approach corresponds to what Hirt et al. (2021) call citation tracking, which expands the coverage of systematic reviews in dispersed research areas with sometimes heterogeneous terminology. The procedure aimed to ensure that publications whose relevance stemmed from intertextual relationships and not simply from keyword matching were included. This option mitigates the biases inherent in relying solely on automated searches.

Figure 1: Citation chaining methodology design



Source: authors' own elaboration.

The final corpus yielded 148 articles. Although Google Scholar deliberately encompasses formats beyond the peer-reviewed article, the corpus is strongly dominated by journal articles, which account for 128 of the 148 items. The remaining items comprise eight conference proceedings, four reports, three book chapters, three preprints, and two working papers. The complete list of the 148 texts, with their bibliographic reference, document type, and DOI, is provided as an appendix to this article. This distribution is itself informative since the preponderance of indexed journal articles reflects where the literature on the effectiveness of fact-checking is currently published and validated, a point we take up in discussing the scope of the search below. These were subjected to a qualitative reading aimed at extracting information about objectives, methods, and results. The subsequent organization was not limited to frequency counting but aimed to delineate convergences and tensions between the studies. We sought to relate the methodological design, thematic domains, and patterns of results. As Pfänder and Altay (2025) argue, the value of this type of review lies less in nominal exhaustiveness and more in the ability to reconstruct the field as a system of relationships. By grouping research to highlight chains of influence, critical responses, and theoretical dependencies, it becomes possible to observe the areas of greatest empirical density as well as areas of uncertainty. The adopted scheme enabled the development of an interpretative framework that dialogues with recent literature on the effects of fact-checking and its variability across contexts (Porter et al., 2021).

The qualitative analysis followed an interpretative rather than a strictly protocolised coding procedure. Each text was read in full and characterised according to a small set of recurrent dimensions, namely the type of effect under examination, the level at which that effect was observed and the family of conditioning factors at stake. These dimensions were not imposed a priori but stabilised through successive readings of the corpus, and they correspond to the cognitive, political, technological, and contextual factors that organise the results section. The categorisation was conducted by the authors through close reading and consensual discussion of ambiguous cases, without recourse to dedicated qualitative-analysis software, in keeping with the interpretative and relational purpose of the review. Where a study addressed prevention or non-professional correction rather than newsroom verification, it was retained and read as a complementary expression of the same sociotechnical problem, consistent with the inclusive understanding of fact-checking set out in the introduction.

A note on the linguistic scope of the search is nonetheless warranted, since the descriptors were formulated in English. We do not consider that this introduced a bias towards Anglophone literature, for two reasons. First, Google Scholar performs a degree of cross-language matching and indexes works whose metadata are in English regardless of their language of publication, which extends its reach to journals indexed internationally irrespective of language. Second, and more decisively, every search expression included the terms "fact-checking" and "fact-check", which in Portuguese, Spanish, and French scholarship are typically left untranslated or coexist with vernacular equivalents. The predominance of internationally indexed journals in the corpus should therefore be read as a feature of where this literature is currently published and validated, rather than as an artefact of the search language. Consistent with this, the review does not claim the exhaustiveness of a systematic literature review: it is a broad-scope, qualitative and reflexive review, and the corpus retrieved corresponds to that purpose.

The four families of conditioning factors that organise the analysis of the second research question, namely the cognitive, political, technological, and contextual, were not adopted as an external grid imposed on the corpus but emerged from the recurrent dimensions along which the studies themselves account for variation in effectiveness, and they can be briefly grounded before the results are presented. Cognitive factors concern the individual processing of corrective information and the conditions of its retention: the literature on motivated reasoning and on the continued influence of misinformation shows that acceptance and durability of a correction depend on processing style, prior knowledge, and the memory dynamics that govern belief regression (Lewandowsky, Ecker & Cook, 2017; Swire-Thompson et al., 2023). Political factors concern partisanship and identity: the strength of prior political attitudes conditions both exposure to fact-checks and their evaluation, so that the same correction may be received as informative or as hostile depending on its alignment with the recipient's commitments (Nyhan & Reifler, 2010; Shin & Thorson, 2017). Technological factors concern the design of the correction and the platform mechanics through which it circulates: the format of the message, the architecture of labels, and the networked and crowdsourced structures of diffusion shape whether a correction reaches its audience and is accepted (Margolin et al., 2018; Park et al., 2021). Contextual factors concern the wider informational environment and the temporal logic of intervention, encompassing the conditions under which correction operates and the shift from reactive verification towards preventive strategies such as inoculation and prebunking (Lewandowsky & van der Linden, 2021). These four dimensions are analytically distinct but empirically entangled, and several studies in the corpus speak to more than one of them simultaneously; the structure of the results section reflects this scaffold while acknowledging that the boundaries between factors are porous.

### 3. Results

Our initial review of the articles collected was able to identify different trends and research contributions related to the study of fact-checking. Overall, the analysis allowed to highlight changes over the years in the perception of effectiveness, limitations, and solutions presented (Table 1). This periodisation responds directly to the third research question (RQ3) and is presented first because it frames the analysis of the remaining two. We then analyze the extent to which empirical research supports the effectiveness of fact-checking in correcting falsehoods (RQ1). Finally, we identify the cognitive, political, technological, and contextual factors that condition both the success and limitations of fact-checking (RQ2).

Table 1: Main research perspectives on fact-checking effectiveness

|                                                            | <b>Trend</b>                                                                                                       | <b>References</b>                                                                                                                                                                                                                                                                                            |
|------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Skeptical perspective</b>                               | Pessimistic outlook. Fact-checking limited, vulnerable to identity and cognitive biases.                           | Nyhan and Reifler, 2010; Kraft et al. 2015; Thorson, 2015; Fazio et al. 2015; Uscinski, 2015; Jarman, 2016; Shin and Thorson, 2017; Westlund et al. 2024                                                                                                                                                     |
| <b>Reevaluation</b>                                        | Paradigm shift. More evidence of effectiveness, albeit limited. Backfire effect is relative.                       | Wood and Porter, 2019; Nyhan et al. 2019; Ecker et al. 2020; Porter and Wood, 2021; Amazeen et al. 2018; Pennycook and Rand, 2019; Walter et al. 2020; Bor et al. 2021; Lee et al. 2022; Chae et al. 2024; Koch et al. 2023; Boukes and Hameleers, 2023; Bachmann and Valenzuela, 2023; Tanihara et al. 2023 |
| <b>Innovation – new formats, design and AI inoculation</b> | Design research, formats (humour, satire, visual), preventive and systemic turn (AI, inoculation, platform design) | Amazeen et al. 2018; Lee et al. 2022; Fazio et al. 2023; Boukes and Hameleers, 2023; Koch et al. 2023; Liu et al. 2025; Moon and Kahlor, 2025; Graham and Porter, 2025; Traberg et al. 2023; Roozenbeek et al. 2020; Lewandowsky & van der Linden, 2021                                                      |

Source: authors' own elaboration

Table 1 summarizes our findings. The review identified a "skeptical perspective" of studies that emphasize the limits of fact-checking. Several research from this period conveys a more pessimistic view, focusing on how political or ideological orientation shapes the consumption, acceptance, and diffusion of fact-checking (Nyhan and Reifler, 2010; Kraft et al. 2015; Thorson, 2015; Fazio et al. 2015; Uscinski, 2015). These studies present unfavorable results regarding the effectiveness of fact-checking in correcting false beliefs, particularly given how this effectiveness can be conditioned by partisan biases. Still, it's important to note that other studies published after 2015 (e.g., Jarman, 2016; Shin and Thorson, 2017; Westlund et al. 2024) demonstrate the limitations of fact-checking, presenting its effectiveness as conditioned by technological, social, cultural, or political factors.

At a second point in time (Table 1), fact-checking appears to be undergoing "reevaluation" as a journalistic practice that combats disinformation. For example, the backfire effect (see Swire-Thompson et al. 2020; Nyhan and Reifler, 2010), also known as the boomerang effect (Hart & Nisbet, 2012), is relativized, and evidence of its rarity is presented when testing the effectiveness of fact-checking (Wood and Porter, 2019; Nyhan et al. 2019; Ecker et al. 2020; Porter and Wood, 2021). This represents a paradigm shift regarding the effect of fact-checking, with more evidence of its effectiveness, albeit limited (e.g., Amazeen et al. 2018; Pennycook and Rand, 2019; Walter et al. 2020; Bor et al. 2021; Lee et al. 2022; Chae et al. 2024; Koch et al. 2023; Boukes and Hameleers, 2023; Bachmann and Valenzuela, 2023; Tanihara et al. 2023). Along with the relative backfire effect on fact-checking consumption and greater evidence of its effectiveness, we

identified studies that focus on "innovation," considering new formats and design strategies applied to fact-checking (Amazeen et al. 2018; Lee et al. 2022; Fazio et al. 2023; Boukes and Hameleers, 2023; Koch et al. 2023).

In a more recent period of our sample, the presence of more studies focused on artificial intelligence, inoculation, and active search is notable (Liu et al. 2025; Moon and Kahlor, 2025; Graham and Porter, 2025; Traberg et al. 2023; Roozenbeek et al. 2020). For example, Moon and Kahlor (2025) investigated how non-human source labels can mitigate the politically motivated reasoning that compromises the effectiveness of fact-checking. This study demonstrated that labeling fact-checks as produced by AI can be an effective strategy for reducing partisan resistance (see also Moon et al., 2023), reinforcing the need to consider source types and label design in fact-checking interventions, especially on politicized science topics.

### 3.1. The effectiveness of fact-checking and related facts

The corpus analyzed was divided into three categories: 1) studies that criticize fact-checking (skeptical or negative view, such as Uscinski, 2015; Arcos et al. 2022; Kanthawala and Maddox, 2022; Pretus et al. 2022); 2) studies that recognize the effect of fact-checking, but with limitations of a cognitive nature (Ecker et al. 2020; Swire-Thompson et al. 2023; Martel et al. 2021), contextual/political (Painter and Fernandes, 2024; Bachmann and Valenzuela, 2023; Walter et al. 2021), format/design (Amazeen et al. 2018; Fazio et al. 2023), reach/consumption (Kyriakidou et al. 2023; Baptista et al. 2023; Poulakidakos et al. 2024) and 3) studies that show, with positive evidence, that fact-checking is effective, with moderate to robust effects and without widespread backfire (e.g., Wood and Porter, 2019; Nyhan et al. 2019; Wintersieck, 2017; Chae et al. 2024; Porter and Wood, 2021; Amazeen et al. 2018; van Erkel et al. 2024).

Nevertheless, in general, most studies demonstrate that fact-checking reduces false beliefs, albeit with limitations such as short persistence, limited reach, contextual barriers, identity-driven effects, source trust, and dependent on digital and media literacy (Q1). Answering the second research question (Q2), our analysis identifies a set of factors that appear to influence the effectiveness of fact-checking, including cognitive styles, political identities, levels of media literacy, technological resources, and platform dynamics. Thus, more specifically, the analysis of our corpus enabled us to identify several research patterns (see Table 2).

Table 2: A literature synthesis of fact-checking effectiveness

| Topic / Category                            | Trend or Pattern                                                                                                                                        | References                                                                                                   |
|---------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| <b>Fact-checking criticism</b>              | Skeptic or negative Outlook; questioning reach, bias, limited effectiveness or format irrelevance in polarized contexts                                 | Uscinski (2015); Arcos et al. (2022); Kanthawala & Maddox (2022); Pretus et al. (2022)                       |
| <b>Cognitive limitations</b>                | Effectiveness dependent on memory, processing Styles and motivated bias; short corrections improve retention; forgetting generates regression of belief | Ecker et al. (2020); Swire-Thompson et al. (2023); Martel et al. (2021)                                      |
| <b>Contextual/ Political limitations</b>    | Political context and strong identity involvement reduce corrective effect; beliefs used to reinforce previous positions                                | Painter & Fernandes (2024); Bachmann & Valenzuela (2023); Walter et al. (2021); Tulin et al. (2025)          |
| <b>Partisanship / Motivated bias</b>        | Partisan identities shape acceptance and sharing; there may be resistance to corrections (backfire is rare)                                             | Nyhan & Reifler (2010); Kraft et al. (2015); Shin & Thorson (2017); Jarman (2016); Bodishtianu et al. (2025) |
| <b>Counterevidence to partisanship bias</b> | Effective corrections regardless of political affiliation; no evidence of widespread backfire                                                           | Wood & Porter (2019); Nyhan et al. (2019); Chae et al. (2024); Wintersieck (2017); Baptista et al. (2023)    |
| <b>Ideological asymmetry</b>                | Conservatives show greater resistance, lower performance, and lower acceptance of fact-checking                                                         | Garrett & Bond (2021); Lyons et al. (2020); Jia & Lee (2024); Shin & Thorson (2017); Aruguete et al. (2025)  |

|                                     |                                                                                                                                                                           |                                                                                                                                                                                                    |
|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Attitudes, support and reach</b> | Low familiarity and reach, but high stated favorability; greater use among young and more educated people                                                                 | Lyons et al. (2020); Baptista et al. (2023); Kyriakidou et al. (2023); Saldaña & Santos (2023); Bode & Vraga (2021); Nyhan & Reifler (2016); Robertson et al. (2020); Godel et al. (2022)          |
| <b>Cognitive styles</b>             | Analytical/reflective thinking and empathy increase acceptance of corrections; mediated by literacy and partisan bias                                                     | Painter & Fernandes (2024); Liu et al. (2025); Martel et al. (2021); Lee & Ramazan (2021)                                                                                                          |
| <b>Literacy and schooling</b>       | Information and media literacy predict consumption and effectiveness; low levels may generate adverse effects                                                             | Tanihara et al. (2023); Bor et al. (2021); Lee & Ramazan (2021); Robertson et al. (2020); Bode & Vraga (2021); Baptista et al. (2023); Poulakidakos et al. (2024)                                  |
| <b>Credibility and sources</b>      | Perceived credibility mediates effectiveness; individual trust moderates; institutional source matters less than content                                                  | Liu et al. (2025); Moon & Kahlor (2025); van Erkel et al. (2024)                                                                                                                                   |
| <b>Format, design and styles</b>    | Short corrections with explanation and repetition increase retention; labels, scales, and humor have mixed effects; headline format is irrelevant, satirical can polarize | Ecker et al. (2020); Amazeen et al. (2018); Walter et al. (2020); Park et al. (2021); Lee et al. (2022); Fazio et al. (2023); Boukes & Hameleers (2023); Autry & Duarte (2021); Wintersieck (2017) |
| <b>Networks and social dynamics</b> | Corrections more often accepted among acquaintances; crowdsourcing can be effective, but remains vulnerable to biases; professional checkers are needed                   | Margolin et al. (2018); Pennycook & Rand (2019); Allen et al. (2022); Draws et al. (2022); Dierickx et al. (2023); Godel et al. (2022)                                                             |
| <b>Innovation and inoculation</b>   | Preventive strategies (prebunking/inoculation) prove effective and scalable, complementing reactive fact-checking                                                         | Lewandowsky & van der Linden (2021); Traberg et al. (2023, 2025); Roozenbeek et al. (2020)                                                                                                         |

Source: authors' own elaboration

### 3.2. Political and partisan factors

The first family of conditioning factors is political. Partisan identity and motivated reasoning shape who reads corrections, how they are evaluated, and whether they translate into belief change, and the evidence here is divided between studies that find partisanship constrains correction and studies that find correction operates largely irrespective of political affiliation.

Regarding the influence of partisanship and the limitations it imposes on the effectiveness of fact-checking in correcting false beliefs, Shin and Thorson (2017) identified a biased tendency toward fact-checking sharing, particularly motivated by pre-existing political orientations, whether liberal or conservative. Jarman (2016) also noted that party affiliation affects the evaluation of political fact-checking. Motivated bias persists. Fact-checking may contribute to reducing beliefs; however, it does not appear to promote opinion convergence, as the distance between political groups remains and may even increase after criticism. Similarly, Tulin et al. (2025) highlighted the influence of fact-checking in the political context related to the Ukrainian war, in which citizens with strong attitudes, especially pro-Ukraine, are more likely to read fact-checking content, not necessarily to change their opinion but to reinforce their position. Nyhan and Reifler (2010) and Kraft et al. (2015) emphasize the idea that corrections and the selection of fact-checking content are intertwined with partisan identities that offer resistance and may even reinforce errors. Bodishtianu et al. (2025) note that while access to fact-checking increases the accuracy of news assessments by more than 40%, individuals tend to check more when the content challenges their prior beliefs, especially in the case of dissonant political views, which corroborates the theory of motivated political reasoning.

On the other hand, other studies (Wood and Porter, 2019; Chae et al. 2024; Baptista et al. 2023; Wintersieck, 2017; Nyhan et al. 2019) found no significant evidence of the impact of partisanship on the accuracy of the assessment of corrected messages. For example, Chae et al. (2024) observed in all experiments that fact-checks significantly reduced agreement and belief in false information, with no

evidence of backfire. Furthermore, regarding real-time fact-checking during televised debates, Wintersieck (2017) found high persuasive power and partisan symmetry regarding the effect of accurate messages. Also in an analysis of Trump and Clinton supporters, Nyhan et al. (2019) showed that both groups, when confronted with corrections, significantly reduced their misperceptions, contradicting the hypothesis of resistant motivation (backfire).

Nevertheless, regarding susceptibility, assimilation, or acceptance of fact-checking, some studies seem to indicate ideological asymmetry. That is, people who are more right-leaning and conservative have a more persistent ideological effect, leading them to perform worse (Garrett and Bond, 2021; Lyons et al. 2020; Jia and Lee, 2024; Shin and Thorson, 2017; Aruguete et al. 2025).

From another perspective, regarding attitudes and support expressed for fact-checking (Table 2), our results seem to highlight a limitation of the practice related to its reach and public engagement (Lyons et al. 2020; Baptista et al. 2023; Kyriakidou et al. 2023), which are dependent on factors such as age and education, with young and more educated people showing higher levels (Baptista et al. 2023; Saldaña and Santos, 2023; Bode and Vraga, 2021; Kyriakidou et al. 2023). Furthermore, the level of familiarity, awareness, and reach of fact-checking practices tends to be higher when there is an increase in citizens' political interest and knowledge (Lyons et al. 2020; Robertson et al. 2020; Nyhan and Reifler, 2016; Godel et al. 2022). Even so, although the results indicate a weak reach of fact-checking, judging by levels of familiarity with the practice, several studies show high levels of favorability (e.g., Lyons et al. 2020; Baptista et al. 2023; Kyriakidou et al. 2023).

### 3.3. Cognitive factors

A second family is cognitive. Here effectiveness depends on how individuals process information, on the memory dynamics that govern whether a correction endures, and on levels of media and information literacy, with the credibility attributed to the source operating as a further mediating condition.

Regarding how different cognitive styles process information, studies show that more analytical, calculating, or reflective information processing tends to be more favorable to correcting false beliefs (Painter and Fernandes, 2024; Liu et al. 2025; Martel et al., 2021; Lee & Ramazan, 2021). Another study, also from a psychological perspective, showed that threat perception and empathy are positively and significantly related to the intention to verify information and a greater responsibility to confirm and correct false content. Notably, cognitive styles can be mediated or influenced by other factors, such as the level of media literacy (Lee and Ramazan, 2021), perceived credibility (Liu et al. 2025), and partisan bias (Painter and Fernandes, 2024). Continuing to analyze our results, we also identified a pattern related to the effect of literacy (information, media) and education on fact-checking consumption (Table 2). Indeed, we observed that literacy and education predict awareness, use, and correction (Tanihara et al. 2023; Bor et al. 2021; Lee and Ramazan, 2021; Robertson et al. 2020; Bode and Vraga, 2021; Baptista et al. 2023; Poulakidakos et al. 2024), while low literacy levels can generate adverse effects (Tanihara et al. 2023). However, individuals with low literacy levels were more likely to revert to "correct" after correction (Tanihara et al. 2023).

Regarding the importance of trust, credibility, or source type, studies indicate that perceived credibility mediates effectiveness, with the source mattering less than the content, but individual trust moderates it (Liu et al. 2025; Moon and Kahlor, 2025; van Erkel et al. 2024).

### 3.4. Technological and format factors

A third family is technological, in the broad sense of the design and the platform mechanics through which corrections are formatted and circulated. This encompasses the form of the correction itself, from explanatory rebuttals and veracity scales to label design, and the networked and crowdsourced dynamics that determine how corrections travel and who accepts them.

These aspects lead our analysis to consider the effect of new formats, designs, or correction styles (Table 2). Ecker et al. (2020) conducted an experiment using short rebuttals of fewer than 140 characters, which included a brief explanatory component rather than merely a false label. The study showed that condensed rebuttals outperform simple "false" labels in the long run and improve critical reasoning. Furthermore, Ecker et al. (2020) showed that repetition can aid corrections, finding that prior exposure to the statement strengthened the effect of rebuttals in reducing false beliefs. Other studies, also focused on innovation in fact-checking, have found that veracity scales increase the effectiveness of fact-checking (Amazeen et al. 2018; Walter et al. 2020). However, Park et al. (2021) showed that label design can have distinct influences, affecting belief accuracy and content circulation. Therefore, it is important to design labels that minimize uncertainty aversion and mitigate the resistance of the most skeptical (Park et al. 2021). In the same field, Lee et al. (2022) observed that simple fact-check labels reduce beliefs, but with limitations, having no impact on reach or engagement generated and that these effects can be conditioned by the topic or context.

Still on the topic of research into new formats or designs, it is worth highlighting the study by Fazio et al. (2023), which sought to understand to what extent the headline format and the number of logos could affect the effectiveness of belief reduction. They observed that fact-checking has a corrective effectiveness; however, the headline format is irrelevant, and logos do not increase credibility or effectiveness. In another approach, Boukes and Hameleers (2023) showed that both regular and satirical fact-checks correct beliefs. However, satirical fact-checking can have an adverse effect, contributing to increased polarization. Wintersieck (2017) demonstrated the persuasive role of instant fact-checking, and Autry and Duarte (2021) showed that corrections formulated with negation (e.g., "it's not red") can produce a familiarity backfire and do not have the desired effect of fact-checking. Other studies in our corpus have focused on the relationship between network structures and moments of interaction on online platforms (Pennycook and Rand, 2019; Margolin et al. 2018; Allen et al. 2022; Draws et al. 2022; Dierickx et al. 2023; Godel et al. 2022). For example, Margolin et al. (2018) examined the social relationship between the corrector and the correctee, showing that there is a shared audience and that the acceptance rate of fact-checking is higher among friends than among strangers. Also in terms of crowdsourcing, several studies have shown positive effects. These assessments are not independent of cognitive and partisan biases, which reinforces the need for professional fact-checkers (Godel et al. 2022; Pennycook and Rand, 2019; Allen et al. 2022; Draws et al. 2022).

### 3.5. Contextual factors and prevention

A fourth family is contextual, concerning the wider conditions under which fact-checking operates and the shift from reactive correction towards prevention. Inoculation and prebunking belong here: they act before exposure rather than after it, and they exemplify the sociotechnical extension of fact-checking beyond newsroom verification that frames this review.

In terms of prevention, several studies view disinformation as a virus, thus focusing on developing a kind of "vaccine," testing the inoculation theory (Lewandowsky & van der Linden, 2021; Traber et al. 2023; Roozenbeek et al. 2020). While traditional fact-checking acts reactively, seeking to correct false information after the fact, inoculation acts preventively, helping the public understand how the world of disinformation works and demystifying its strategies (Lewandowsky and van der Linden, 2021). Inoculation appears to be a promising initiative to combat disinformation and sometimes surpasses reactive corrections in effectiveness and reach (Lewandowsky and van der Linden, 2021; Traber et al. 2023). Traber et al. (2023) also provide an integrated model of effectiveness, showing that pre-bunking complements and anticipates reactive interventions, offering guidelines for message and platform design. Roozenbeek et al. (2020) test the effectiveness of an online game, based on the inoculation theory, to "pre-bunk" users against six common disinformation techniques, evaluating this effect in five languages and four different cultural contexts. The study confirmed that proactive and playful inoculation-based interventions can be a scalable and culturally universal strategy to reinforce cognitive resistance against disinformation tactics (Roozenbeek et al. 2020).

#### **4. Discussion and Conclusion**

This review aimed to address three central research questions regarding the effectiveness of fact-checking the conditions under which it operates, and the evolution of research on the practice. In response to the first question, the literature showed a growing consensus that fact-checking can reduce false beliefs, although its impact appears neither uniform nor unconditional. Regarding the second question, the results demonstrate that effectiveness is systematically conditioned by several factors, namely cognitive, political, technological, and contextual. These factors influence how corrections are processed, considered reliable, and disseminated in contemporary informational environments. In response to the third question, the review identified three successive phases of scholarly understanding, moving from an initial phase of skepticism about corrective effects, through a reassessment that established a conditional but real effectiveness and relativised the backfire effect, towards a recent phase centred on explanatory formats, prevention, and automation. This trajectory shows the field progressively abandoning the binary question of whether fact-checking works, in favour of conditional and preventive paradigms, a movement that grounds the interpretative argument developed below.

In general, and in response to our research question, we identified an emerging consensus regarding the effectiveness of fact-checking, presenting it as a practice, not exclusive to journalists, that tends to be effective in correcting erroneous beliefs, although strongly conditioned by individual factors (such as different cognitive styles of processing information, partisan political motivations, or levels of media and information literacy). This formulation is deliberate and consistent with the scope set out at the outset: it does not dilute the journalistic identity of fact-checking but locates that identity within a broader sociotechnical field. Journalism remains the centre of gravity and the normative reference of the practice, yet the corrective effort it embodies is increasingly shared with computational, crowdsourced, and preventive interventions that operate alongside professional verification. To say that effectiveness is not exclusive to journalists is therefore not to deny the journalistic ethos invoked earlier, but to recognise that the conditions of effectiveness identified in this review hold across that wider field rather than within the newsroom alone.

It is important to emphasize that this review is a new contribution, especially with regard to the interpretation of the apparent inconsistencies reported in the literature on this topic. Therefore, the literature analyzed, instead of reflecting contradictory evidence on the effectiveness of fact-checking, presents differences that appear to be the result of studies conducted under different analytical conditions. Specifically, it is observed that studies carried out in more polarized political contexts or focusing on the persistence of beliefs in the long term, more frequently report more limited corrective effects, while other research, of an experimental nature or that measures immediate cognitive outcomes, tends to identify more robust effects (Walter et al., 2020; Ecker et al., 2020; Porter & Wood, 2021). In some ways, this seems to suggest that the effectiveness of fact-checking should not be understood as an intrinsic property of the practice itself, but rather as the result of a set of interactions between conditions, not only cognitive, but also political or technological (Lewandowsky et al., 2017; Swire-Thompson & Lazer, 2020). This interpretative perspective helps to reconcile findings that have often been presented as inconsistent and reinforces the need to analyze these dimensions as interdependent, rather than as isolated explanatory factors.

The results highlight the need to promote media and information literacy programs from an early age, while also investing in measures (triggers or stimuli) that foster critical thinking. It is therefore important to approach fact-checking as a practice of multidimensional complexity—not only by examining how it functions, but more critically, by considering for whom, under what conditions, and to what effect it operates.

Furthermore, our literature review showed that several studies have sought to innovate in the design and format of fact-checking; and found that using simple labels (Lee et al. 2022), veracity scales (Amazeen et al. 2018; Walter et al. 2020), short and explanatory rebuttals (Ecker et al. 2020), and more visual and interactive formats (Young et al. 2018; Wintersieck, 2017) can contribute to increasing the effectiveness and impact of fact-checking.

On the other hand, we identified limitations to fact-checking reach and engagement correlated with low levels of familiarity and use (Kyriakidou et al. 2023; Baptista et al. 2023). We therefore contend that it is necessary to develop strategies that foster a more active demand for fact-checks, particularly through the promotion of stronger community engagement within networks. In this sense, we believe that fact-checking should not be understood as an isolated corrective intervention, but as part of a broader set of professional routines, editorial projects, and audience-oriented strategies.

Our results also reinforce the importance of investing in prevention, with a more proactive approach to combating disinformation. Applying inoculation strategies appears to reduce the cognitive resistance that persists, particularly with motivated political reasoning.

In conclusion, our study proposes an integrated interpretation of evidence that has often been examined in isolation regarding the effectiveness of fact-checking. In fact, our research showed that it is not looking for a miracle solution or a perfect formula to combat misinformation, but presents results suggesting that fact-checking should be understood as a multidimensional socio-technical intervention, in which effectiveness has been shown to be more dependent on an interaction between psychological, political, technological, and contextual factors. Furthermore, it is important to emphasize that we believe future research should go beyond the question of whether fact-checking works universally. That is, in the future, it is necessary to investigate how these dimensions interact in different political or media environments. It is important to reinforce the focus on more longitudinal analyses, on the cumulative effects of repeated exposure to

corrections, and on the growing role of artificial intelligence in supporting not only reactive fact-checking, but also more preventive approaches, such as pre-debunking and inoculation.

### Acknowledgements / Funding

This research was supported by Foundation for Science and Technology (FCT) under Grant Ref. 2023.12581.PEX.

### Conflict of interest

The author(s) declare no conflict of interest.

### Ethical statement

This study was conducted in accordance with the principles of scientific research and did not require additional ethics committee approval.

### Data availability

The data supporting the findings of this study are available upon request.

### Declaration of AI usage

Generative AI tools were used for translation and formatting. Model GPT 5.5.

### Author contributions

|                            | João Pedro Baptista | Francisco Conrado Filho | Anabela Gradim |
|----------------------------|---------------------|-------------------------|----------------|
| Conceptualization          | X                   |                         |                |
| Data curation              | X                   | X                       |                |
| Formal analysis            | X                   |                         |                |
| Funding acquisition        | X                   |                         |                |
| Investigation              | X                   | X                       |                |
| Methodology                | X                   |                         |                |
| Project administration     | X                   |                         |                |
| Resources                  | X                   | X                       |                |
| Software                   | X                   |                         |                |
| Supervision                | X                   |                         |                |
| Validation                 |                     | X                       | X              |
| Writing - original draft   | X                   | X                       |                |
| Writing - review & editing |                     | X                       | X              |

### References

- Allen, J. A., Martel, C., & Rand, D. G. (2022). Birds of a Feather Don't Fact-Check Each Other: Partisanship and the Evaluation of News in Twitter's Birdwatch Crowdsourced Fact-Checking Program. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–19). ACM. <https://doi.org/10.1145/3491102.3502040>
- Amazeen, M. A. (2020). Journalistic interventions: The structural factors affecting the global emergence of fact-checking. *Journalism*, 21(1), 95-111. <https://doi.org/10.1177/1464884917730217>
- Amazeen, M. A., Thorson, E., Muddiman, A., & Graves, L. (2018). Correcting Political and Consumer Misperceptions: The Effectiveness and Effects of Rating Scale Versus Contextual Correction Formats.

- Journalism & Mass Communication Quarterly*, 95(1), 28–48. <https://doi.org/10.1177/1077699016678186>
- Arcos, R., Gertrudix, M., Arribas, C., & Cardarilli, M. (2022). Responses to digital disinformation as part of hybrid threats: a systematic review on the effects of disinformation and the effectiveness of fact-checking/debunking. *Open Research Europe*, 2(8). <https://doi.org/10.12688/openreseurope.14088.1>
- Aruguete, N., Calvo, E., & Ventura, T. (2025). The fact-checking dilemma: Fact-checking increases the reputation of the fact-checker but creates perceptions of ideological bias. *Research & Politics*, 12(1). <https://doi.org/10.1177/20531680251323120>
- Autry, K. S., & Duarte, S. E. (2021). *Correcting the unknown: Negated corrections may increase belief in misinformation*. *Applied Cognitive Psychology*, 35(4), 960–975. <https://doi.org/10.1002/acp.3823>
- Bachmann, I., & Valenzuela, S. (2023). Studying the Downstream Effects of Fact-Checking on Social Media: Experiments on Correction Formats, Belief Accuracy, and Media Trust. *Social Media + Society*, 9(2). <https://doi.org/10.1177/20563051231179694>
- Baptista, J. P., Gradim, A., Loureiro, M., & Ribeiro, F. (2023). Fact-checking: uma prática recente em Portugal? Análise da percepção da audiência. *Anuario Electrónico de Estudios en Comunicación Social "Disertaciones"*, 16(1), 1–28. <https://doi.org/10.12804/revistas.urosario.edu.co/disertaciones/a.12426>
- Bode, L., & Vraga, E. K. (2021). Correction Experiences on Social Media During COVID-19. *Social Media + Society*, 7(2). <https://doi.org/10.1177/20563051211008829>
- Bodishtianu, V., Gaozhao, D., & Zhang, P. (2025). The Motivation and Consequence of Fact-Checking Behavior: An Experimental Study. *PLoS ONE*, 20(5), e0323105. <https://doi.org/10.1371/journal.pone.0323105>
- Bor, A., Osmundsen, M., Rasmussen, S. H. R., Bechmann, A., & Petersen, M. B. (2021). "Fact-checking" videos reduce belief in, but not the sharing of fake news on Twitter. [Working Paper]. Aarhus University
- Boukes, M., & Hameleers, M. (2023). Fighting lies with facts or humor: Comparing the effectiveness of satirical and regular fact-checks in response to misinformation and disinformation. *Communication Monographs*, 90(1), 69–91. <https://doi.org/10.1080/03637751.2022.2097284>
- Chae, J. H., Lee, S. Y., & Song, H. (2024). Perceiving as Biased but Nevertheless Persuaded? Effects of Fact-Checking News Delivered by Partisan Media. *Political Psychology*, 45(1), 69–89. <https://doi.org/10.1111/pops.12914>
- Cheruiyot, D., & Ferrer-Conill, R. (2018). "Fact-Checking Africa": Epistemologies, data and the expansion of journalistic discourse. *Digital Journalism*, 6(8), 964–975. <https://doi.org/10.1080/21670811.2018.1493940>
- Delgado-López-Cózar, E., & Repiso-Caballero, R. (2013). El impacto de las revistas de comunicación: comparando Google Scholar Metrics, Web of Science y Scopus. *Comunicar*, 41, 45–52. <https://doi.org/10.3916/C41-2013-04>
- Dierickx, L., Lindén, C.-G., & Opdahl, A. L. (2023). Automated Fact-Checking to Support Professional Practices: Systematic Literature Review and Meta-Analysis. *International Journal of Communication*, 17, 5170–5190.
- Draws, T., La Barbera, D., Soprano, M., Roitero, K., Ceolin, D., Checchio, A., & Mizzaro, S. (2022, June). The effects of crowd worker biases in fact-checking tasks. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2114–2124). Association for Computing Machinery. <https://doi.org/10.1145/3531146.3534629>
- Ecker, U. K. H., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology*, 111(1), 36–54. <https://doi.org/10.1111/bjop.12383>
- Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993–1002. <https://doi.org/10.1037/xge0000098>

- Fazio, L. K., Hong, M. K., & Pillai, R. M. (2023). Combatting Rumors around the French Election: The Memorability and Effectiveness of Fact-Checking Articles. *Cognitive Research: Principles and Implications*, 8(44). <https://doi.org/10.1186/s41235-023-00500-2>
- Garrett, R. K., & Bond, R. M. (2021). Conservatives' susceptibility to political misperceptions. *Science Advances*, 7(23), eabf1234. <https://doi.org/10.1126/sciadv.abf1234>
- Godel, W., Sanderson, Z., Aslett, K., Nagler, J., Bonneau, R., Persily, N., & Tucker, J. A. (2022). Moderating with the Mob: Evaluating the Efficacy of Real-Time Crowdsourced Fact-Checking. *Journal of Online Trust and Safety*, 1(1), 1–36. <https://doi.org/10.54501/jots.v1i1.15>
- Graham, M. H., & Porter, E. V. (2025). Increasing Demand for Fact-Checking. *Political Communication*, 42(2), 325–348. <https://doi.org/10.1080/10584609.2024.2395859>
- Graves, L. (2016). *Deciding what's true: The rise of political fact-checking in American journalism*. Columbia University Press.
- Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*, 118(1), 177–214. <https://doi.org/10.1007/s11192-018-2958-5>
- Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLOS ONE*, 10(9), e0138237. <https://doi.org/10.1371/journal.pone.0138237>
- Harsin, J. (2015). Regimes of posttruth, postpolitics, and attention economies. *Communication, culture & critique*, 8(2), 327–333. <https://doi.org/10.1111/cccr.12097>
- Hart, P. S., & Nisbet, E. C. (2012). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research*, 39(6), 701–723. <https://doi.org/10.1177/0093650211416646>
- Hirt, J., Nordhausen, T., Appenzeller-Herzog, C., & Ewald, H. (2021). Using citation tracking for systematic literature searching-study protocol for a scoping review of methodological studies and a Delphi study. *F1000Research*, 9, 1386. <https://doi.org/10.12688/f1000research.27337.3>
- Jarman, J. W. (2016). Influence of Political Affiliation and Criticism on the Effectiveness of Political Fact-Checking. *Communication Research Reports*, 33(1), 9–15. <https://doi.org/10.1080/08824096.2015.1117436>
- Jia, C., & Lee, T. (2024). Journalistic Interventions Matter: Understanding How Americans Perceive Fact-Checking Labels. *Harvard Kennedy School Misinformation Review*, 5(2). <https://doi.org/10.37016/mr-2020-138>
- Kanthawala, S., & Maddox, J. (2022). Hiding in the Echo Chamber: Fact-Checking Failures and Individual Tactics of Accuracy Determination on WhatsApp in India. *Asian Journal of Communication*, 32(2), 174–191. <https://doi.org/10.1080/01292986.2021.2023594>
- Koch, T. K., Frischlich, L., & Lerner, E. (2023). Effects of fact-checking warning labels and social endorsement cues on climate change fake news credibility and engagement on social media. *Journal of Applied Social Psychology*, 53(5), 495–507. <https://doi.org/10.1111/jasp.12959>
- Kraft, P. W., Lodge, M., & Taber, C. S. (2015). Why People “Don’t Trust the Evidence”: Motivated Reasoning and Scientific Beliefs. *The ANNALS of the American Academy of Political and Social Science*, 658(1), 121–133. <https://doi.org/10.1177/0002716214554758>
- Kyriakidou, M., Cushion, S., Hughes, C., & Morani, M. (2023). Questioning Fact-Checking in the Fight Against Disinformation: An Audience Perspective. *Journalism Practice*, 17(10), 2123–2139. <https://doi.org/10.1080/17512786.2022.2097118>
- Lee, D. K. L., & Ramazan, O. (2021). Fact-Checking of Health Information: The Effect of Media Literacy, Metacognition and Health Information Exposure. *Journal of Health Communication*, 26(7), 491–500. <https://doi.org/10.1080/10810730.2021.1955312>
- Lee, J., Kim, J. W., & Lee, H. Y. (2022). Unlocking Conspiracy Belief Systems: How Fact-Checking Label on Twitter Counters Conspiratorial MMR Vaccine Misinformation. *Health Communication*, 38(9), 1780–1792. <https://doi.org/10.1080/10410236.2022.2031452>

- Lewandowsky, S., & van der Linden, S. (2021). Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology*, 32(2), 348–384. <https://doi.org/10.1080/10463283.2021.1876983>
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the “Post-Truth” Era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353–369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- 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. <https://doi.org/10.1177/1529100612451018>
- Liu, X., Qi, L., Wang, L., & Metzger, M. J. (2025). Checking the Fact-Checkers: The Role of Source Type, Perceived Credibility, and Individual Differences in Fact-Checking Effectiveness. *Communication Research*, 52(6), 719–746. <https://doi.org/10.1177/00936502231206419>
- Lyons, B., Mérola, V., Reifler, J., & Stoeckel, F. (2020). How Politics Shape Views Toward Fact-Checking: Evidence from Six European Countries. *The International Journal of Press/Politics*, 25(3), 469–492. <https://doi.org/10.1177/1940161220921732>
- Margolin, D. B., Hannak, A., & Weber, I. (2018). Political Fact-Checking on Twitter: When Do Corrections Have an Effect? *Political Communication*, 35(2), 196–219. <https://doi.org/10.1080/10584609.2017.1334018>
- Martel, C., Mosleh, M., & Rand, D. G. (2021). You’re Definitely Wrong, Maybe: Correction Style Has Minimal Effect on Corrections of Misinformation Online. *Media and Communication*, 9(1), 120–133. <https://doi.org/10.17645/mac.v9i1.3519>
- McIntyre, L. (2018). *Post-Truth*. Cambridge, MA: MIT Press.
- Moon, W.-K., & Kahlor, L. A. (2025). Fact-Checking in the Age of AI: Reducing Biases with Non-Human Information Sources. *Technology in Society*, 80, 102760. <https://doi.org/10.1016/j.techsoc.2024.102760>
- Moon, W.-K., Chung, M., & Jones-Jang, S. M. (2023). How Can We Fight Partisan Biases in the COVID-19 Pandemic? AI Source Labels on Fact-checking Messages Reduce Motivated Reasoning. *Mass Communication and Society*, 26(4), 646–670. <https://doi.org/10.1080/15205436.2022.2097926>
- Nieminen, S., & Rapeli, L. (2018). Fighting Misperceptions and Doubting Journalists’ Objectivity: A Review of Fact-Checking Literature. *Political Studies Review*, 17(3), 296–309. <https://doi.org/10.1177/1478929918786852>
- Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, 32(2), 303–330. <https://doi.org/10.1007/s11109-010-9112-2>
- Nyhan, B., & Reifler, J. (2015). *Estimating fact-checking’s effects*. Arlington, VA: American Press Institute.
- Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2019). Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability. *Political Behavior*, 42(3), 939–960. <https://doi.org/10.1007/s11109-019-09528-x>
- Painter, D. L., & Fernandes, J. (2024). “The Big Lie”: How Fact-Checking Influences Support for Insurrection. *American Behavioral Scientist*, 68(7), 892–912. <https://doi.org/10.1177/00027642221103179>
- Park, S., Park, J. Y., Chin, H., Kang, J. H., & Cha, M. (2021). *An experimental study to understand user experience and perception bias occurred by fact-checking messages*. In Proceedings of the web conference 2021 (pp. 2769–2780). <https://doi.org/10.1145/3442381.3450121>
- Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 116(7), 2521–2526. <https://doi.org/10.1073/pnas.1806781116>
- Pfänder, J., & Altay, S. (2025). Spotting false news and doubting true news: A systematic review and meta-analysis of news judgements. *Nature Human Behaviour*, 9(4), 688–699. <https://doi.org/10.1038/s41562-024-02086-1>

- 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), e2104235118. <https://doi.org/10.1073/pnas.2104235118>
- Poulakidakos, S., Demertzis, N., & Linardis, A. (2024). Perceptions and attitudes towards fact-checking in Greece: A quantitative audience research. *Science and Society: Journal of Political and Moral Theory*, 43, 34–68.
- Pretus, C., Servin-Barthet, C., Harris, E. A., Brady, W. J., Vilarroya, O., & Van Bavel, J. J. (2022). The role of political devotion in sharing partisan misinformation. *Preprint*. <https://doi.org/10.21203/rs.3.rs-1665189/v1>
- Robertson, C. T., Mourão, R. R., & Thorson, E. K. (2020). Who Uses Fact-Checking Sites? The Impact of Demographics, Political Antecedents, and Media Use on Fact-Checking Site Awareness, Attitudes, and Behavior. *The International Journal of Press/Politics*, 25(2), 217–237. <https://doi.org/10.1177/1940161219898055>
- Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on “inoculation” theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*, 1(2), Article 8. <https://doi.org/10.37016//mr-2020-008>
- Saldaña, M., & Santos, M. (2023). Sherlock-wannabes or when the audience fact-checks: How ideology, education, and alternative media use explain individual fact-checking behaviors in Chile. *Estudios sobre el Mensaje Periodístico*, 29(4), 795–805. <https://doi.org/10.5209/esmp.88097>
- Shin, J., & Thorson, K. (2017). Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media. *Journal of Communication*, 67(2), 200–228. <https://doi.org/10.1111/jcom.12284>
- 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. <https://doi.org/10.1016/j.jarmac.2020.06.006>
- Swire-Thompson, B., & Lazer, D. (2020). Public Health and Online Misinformation: Challenges and Recommendations. *Annual Review of Public Health*, 41, 433–451. <https://doi.org/10.1146/annurev-publhealth-040119-094127>
- Swire-Thompson, B., Dobbs, M., Thomas, A., & DeGutis, J. (2023). Memory failure predicts belief regression after the correction of misinformation. *Cognition*, 230, 105276. <https://doi.org/10.1016/j.cognition.2022.105276>
- Tanihara, T., Yamaguchi, S., & Watanabe, T. (2023). Literacy Is Necessary to Understand Fact-Checking: An Empirical Research Using Survey Experiments. *Journal of Socio-Informatics*, 16(1), 33–46. [https://doi.org/10.14836/jsi.16.1\\_33](https://doi.org/10.14836/jsi.16.1_33)
- Thorson, E. (2015). Belief Echoes: The Persistent Effects of Corrected Misinformation. *Political Communication*. <https://doi.org/10.1080/10584609.2015.1102187>
- Traberg, C. S., Harjani, T., Basol, M., Biddlestone, M., Maertens, R., Roozenbeek, J., & van der Linden, S. (2023). Prebunking against misinformation in the modern digital age. *Managing Infodemics in the 21st Century: Addressing New Public Health Challenges in the Information Ecosystem*, 99–111. [https://doi.org/10.1007/978-3-031-27789-4\\_8](https://doi.org/10.1007/978-3-031-27789-4_8)
- Tulin, M., Hameleers, M., de Vreese, C., Aalberg, T., Corbu, N., Van Erkel, P., Esser, F., Gehle, L., Halagiera, D., Hopmann, D. N., Koc-Michalska, K., Matthes, J., Mihelj, S., Schemer, C., Stetka, V., Strömbäck, J., Terren, L., & Theocharis, Y. (2025). Why do Citizens Choose to Read Fact-Checks in the Context of the Russian War in Ukraine? The Role of Directional and Accuracy Motivations in Nineteen Democracies. *The International Journal of Press/Politics*, 30(3), 679–704. <https://doi.org/10.1177/19401612241233533>
- Uscinski, J. E. (2015). The Epistemology of Fact Checking (Is Still Naïve): Rejoinder to Amazeen. *Critical Review*, 27(2), 243–252. <https://doi.org/10.1080/08913811.2015.1055892>
- Uscinski, J. E., & Butler, R. W. (2013). The Epistemology of Fact Checking. *Critical Review*, 25(2), 162–180. <https://doi.org/10.1080/08913811.2013.843872>

- van Erkel, P. F. A., van Aelst, P., de Vreese, C. H., Hopmann, D. N., Matthes, J., Stanyer, J., & Corbu, N. (2024). When Are Fact-Checks Effective? An Experimental Study on the Inclusion of the Misinformation Source and the Source of Fact-Checks in 16 European Countries. *Mass Communication and Society*, 27(5), 851–876. <https://doi.org/10.1080/15205436.2024.2321542>
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3), 423–441. <https://doi.org/10.1080/03637751.2018.1467564>
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*, 37(3), 350–375. <https://doi.org/10.1080/10584609.2019.1668894>
- Walter, N., Edgerly, S., & Saucier, C. J. (2021). “Trust, then verify”: When and why people fact-check partisan information. *International Journal of Communication*, 15, 4734–4754.
- Westlund, O., Belair-Gagnon, V., Graves, L., Larsen, R., & Steensen, S. (2024). What Is the Problem with Misinformation? Fact-checking as a Sociotechnical and Problem-Solving Practice. *Journalism Studies*, 25(8), 898–918. <https://doi.org/10.1080/1461670X.2024.2357316>
- Wintersieck, A. L. (2017). Debating the Truth: The Impact of Fact-Checking During Electoral Debates. *American Politics Research*, 45(2), 304–331. <https://doi.org/10.1177/1532673X16686555>
- Wood, T., & Porter, E. (2019). The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence. *Political Behavior*, 41(1), 135–163. <https://doi.org/10.1007/s11109-018-9443-y>
- Young, D. G., Jamieson, K. H., Poulsen, S., & Goldring, A. (2018). Fact-Checking Effectiveness as a Function of Format and Tone: Evaluating FactCheck.org and FlackCheck.org. *Journalism & Mass Communication Quarterly*, 95(1), 49–75. <https://doi.org/10.1177/1077699017710453>