

RESEARCH ARTICLE



# Can Fighting Misinformation Have a Negative Spillover Effect? How Warnings for the Threat of Misinformation Can Decrease General News Credibility

Toni G. L. A. van der Meer<sup>a</sup>, Michael Hameleers<sup>a</sup> and Jakob Ohme<sup>b</sup>

<sup>a</sup>Amsterdam School of Communications Research (ASCoR), University of Amsterdam, Amsterdam, The Netherlands; <sup>b</sup>Weizenbaum Institute for the Networked Society, Freie Universität Berlin, Berlin, Germany

## ABSTRACT

In the battle against misinformation, do negative spillover effects of communicative efforts intended to protect audiences from inaccurate information exist? Given the relatively limited prevalence of misinformation in people's news diets, this study explores if the heightened salience of misinformation as a persistent societal threat can have an unintended spillover effect by decreasing the credibility of factually accurate news. Using an experimental design ( $N=1305$ ), we test whether credibility ratings of factually accurate news are subject to exposure to misinformation, corrective information, misinformation warnings, and news media literacy (NML) interventions relativizing the misinformation threat. Findings suggest that efforts like warning about the threat of misinformation can prime general distrust in authentic news, hinting toward a deception bias in the context of fear of misinformation being salient. Next, the successfulness of NML interventions is not straight forward if it comes to avoiding that the salience of misinformation distorts people's credibility accuracy. We conclude that the threats of the misinformation order may not just be remedied by fighting false information, but also by reestablishing trust in legitimate news.

## ARTICLE HISTORY

Received 5 September 2022  
Accepted 27 February 2023

## KEYWORDS

Misinformation; news credibility; news media literacy; corrective information; misinformation warning; deception bias

The spread of misinformation has raised ample concerns regarding the distortion of our (digital) news environment (Pennycook et al. 2021). This threat of false information has potentially altered or even biased, the ways in which audiences process and assess information. For example, the challenging task of deception detection—i.e., detecting the accuracy of fake and real news—can become more salient when browsing news (Luo, Hancock, and Markowitz 2022). Accordingly, a considerable wealth of interventions, ranging from factchecks to warnings about misinformation's omnipresence and inoculation strategies (e.g., Roozenbeek and van der Linden 2019), have been introduced in an attempt to reduce the potential harm of false information (Acerbi, Altay, and

**CONTACT** Toni G. L. A. van der Meer  [g.l.a.vandermeer@uva.nl](mailto:g.l.a.vandermeer@uva.nl)

Supplemental data for this article can be accessed at <https://doi.org/10.1080/1461670X.2023.2187652>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Mercier 2022). The current study takes a next step in exploring whether such well-intended attempts to fight misinformation might incur unintended consequences for audiences' overall perceptions of news, namely the credibility assessment of factual correct information.

The truth-default theory argues that people hold a truth-bias in assessing the credibility of information (Levine 2021). The central premise of this theory is that individuals tend to, by default, presume that what others say is basically honest (Levine 2014). Yet, in the current digital news ecology, where public concern on misinformation abounds (e.g., Newman et al. 2022), the question is whether this default state is viable. The truth-default theory might not hold true for news assessment in a society alarmed by the fast spread of misinformation. Within the context of a so-called post-truth era (Van Aelst et al. 2017), audiences might be more attentive to the potential deception of information, as recently found for social media content (Luo, Hancock, and Markowitz 2022). While this may lead to a sensitization towards misinformation, it may distort the way accuracy is detected and thereby results in an over-estimation of the deception threat and prime general distrust even for authentic and factually correct news.

Although the consequences of misinformation can be far-reaching, some research has shown that such false information comprises less than one percent of people's news diets (Allen et al. 2020; Grinberg et al. 2019) and that the base rate of online exposure to misinformation is very low (see Acerbi, Altay, and Mercier (2022) for an overview). Despite that most societal threats are generally objectively rare (e.g., terrorist attacks, nuclear accidents), the relatively exceptional spread of misinformation might, at times, be overrepresented in public discourse—partly because media attention is biased toward the negative and rare (Soroka, Fournier, and Nir 2019; van der Meer et al. 2019)—and efforts to fight it primarily heighten the salience of misinformation as a threat. When people are generally more exposed to reliable information, nudges to fight misinformation will mainly affect the processing of reliable information, ergo, critical information processing will predominantly be applied to scrutinize reliable information (Altay 2022). The relativization of misinformation's prominence is, however, typically left out of public discussions and pre- and debunking initiatives that focus on detecting false information. The high salience of misinformation as a persistent threat, despite its rarity, might cultivate disproportionate skepticism about all news, including reliable information. It is therefore important to assess whether interventions may spillover, and how this potential side-effect may be avoided by relativizing the threat of misinformation.

Since being able to form accurate beliefs based on credible information is at the heart of a well-functioning democracy (Pennycook, Cannon, and Rand 2018), the present study is set out to assess if the public debate about misinformation and communicative attempts to correct it (such as fact-checking and misinformation warnings) may have wide-ranging consequences for epistemic perceptions about news—i.e., accuracy of information credibility assessment. These communicative efforts can correct misperceptions based on misinformation exposure, yet they might also heighten the salience of misinformation as a persistent societal threat, potentially causing a negative spillover effect by decreasing credibility judgement of factually accurate news. The overall research question, hence, is: Can exposure to communicative attempts to fight misinformation decline credibility of factual news?

With the use of an experimental design, we aim to understand how credibility ratings of factual and legitimate news are subject to prior exposure to (i) misinformation, (ii)

corrective information, (iii) misinformation warnings, and (iv) news media literacy (NML) interventions. The study contributes to existing research, first, by offering empirical evidence that attention to misinformation in the digitized media ecology may have unintended consequences. Second, we offer new insights into the effectiveness of pre-bunking initiatives in light of consolidating and restoring trust in honest and correct information. Accordingly, we respond to the call to develop media literacy interventions that help citizens to critically navigate their information environment amidst an era characterized by widespread distrust (Mihailidis and Viotty 2017).

### ***News Credibility and Accuracy Detection***

We rely on literature about message credibility to study how accurately citizens assess reliable and factually correct news in a post-truth area. Citizens' trust in information sources like the news media is fundamental for a well-informed and well-functioning democracy, as citizens need to base their opinions and disagreements on a consensual truth. Media trust refers to people's overall evaluation of the media and in particular refers to information credibility (Kohring and Matthes 2007). News credibility has therefore been described as the core element of media trust (e.g., Tsfaty and Cappella 2003). It is understood as an encompassing evaluation that combines the assessment of general accuracy in journalistic performance as well as the selection of topics and facts (Kohring and Matthes 2007). The current study is specifically interested in credibility assessment on the level of the message/content of information (Appelman and Sundar 2016).

News credibility is a subjective assessment of individuals. Besides being driven by perceptual message credibility, accuracy detection is another important driver of credibility, as it includes whether such assessments are correct relative to the ground truth of the message (Vrij 2008). Prior work already highlighted that making accurate decisions on credibility is a challenging task (McGrew et al. 2018) and that individuals have difficulties discerning fake from real news. Certain common audience characteristics, such as reflexive open-mindedness (Pennycook and Rand 2020) and lower propensity to engage in analytical reasoning (Pennycook and Rand 2019), are associated with difficulties discerning fake from real news. Due to this difficulty and the potential exhaustion of assessing news credibility for every news article people come across, the shifts in the information ecology may come with biases in credibility assessment.

### ***Truth or Deception Default***

Judgments of information veracity are found to be not without biases. The truth-default theory (Levine 2014) provides a theoretical starting point to understand defects in news credibility assessment. This theory, at its fundamental core, assumes that individuals typically presume that others are honest. This truth-bias has provided a theoretical foundation for understanding how people (inaccurately) process information in interpersonal communication (Gilbert 1991). In communication with others, a passive presumption of honesty tends to, by default, be stronger than the idea of deception (Clare and Levine 2019). Overall, information is thus more likely to be rated as honest than dishonest. As the proportion of true information is assumed to be higher than the proportion

of false information, the truth-default should offer an effective heuristic (Levine 2014). This can be explained by the so-called veracity effect (Levine, Park, and McCornack 1999), which describes that due to the truth-bias, the assessment for truths are more accurate compared to lies. A meta study has concluded that people are truth-biased since they more accurately judge truths as truthful than lies as dishonest (Bond and DePaulo 2006).

In contrast to the truth default, the deception bias (Bond et al. 2005) and interpersonal deception theory (Buller and Burgoon 1996; Burgoon 2015) state that the default assessment of senders' dishonesty occurs at times and that people tend to monitor for deception while being aware of the veracity of information. Truth-default theory does specify that consciously considering the possibility of deceit, rather than the default mechanism of considering information as truthful, can occur in the presence of certain trigger events that prompt suspicion (Levine 2014). Accordingly, deception bias occurs as a product of triggers that facilitate suspicion, such as lack of coherence or consistency with reality and general news content (Luo, Hancock, and Markowitz 2022). Examples of such classes of trigger events are suspected deception motive, dishonest demeanor, logical inconsistencies (between statement or with known facts), and third-party information (Clare and Levine 2019). In our paper, we specifically focus on how the context of misinformation can actively trigger such suspicion by priming the idea of omnipresent falsehoods.

### ***Misinformation and Deception Bias***

In the current news ecology where people are constantly alarmed by misinformation, it is important to revisit the truth bias and how citizens form their judgement on news credibility (Clare and Levine 2019). Here we define misinformation as an umbrella-term that comprises any type of false information or information that can be deemed untrue or incorrect based on relevant expert knowledge and/or empirical evidence (e.g., Vraga and Bode 2020). As a first step, this study assumes that exposure to misinformation has the potential to create misperceptions and distort audiences' truth ratings of news. False information is problematic to the extent that it results in inaccurate beliefs and facilitates decision-making based on misperceptions (Luo, Hancock, and Markowitz 2022; Southwell, Thorson, and Sheble 2018). Related to the processing of incoming information, misinformation studies have primarily focused on how, for example, motivated reasoning and cues in the news environment can drive misperception or perception of false information (e.g., Nyhan and Reifler 2010). More recently, research has documented how respondents were inclined to suspect headlines to be fake news (Luo, Hancock, and Markowitz 2022) and are just as likely to accept fake news as they are to reject true news (Altay, de Araujo, and Mercier 2022; Pennycook et al. 2020).

We ask here how misinformation exposure affects the processing of factually correct news. On the one hand, misinformation might distort people's perception to an extent they might be less likely to believe factual information on the matter. On the other hand, after exposure to misinformation, especially when it is recognized as such, it might trigger citizens to be more aware of potential deception in other news, even if it runs counter to the initial misinformation piece. In other words, we ask whether exposure to misinformation shifts news credibility assessment from a truth bias towards a deception bias. Therefore, we explore if, after exposure to misinformation, citizens are more likely to falsely consider factual information as inaccurate:

**RQ1.** Can exposure to misinformation decrease people's rating of the credibility of factually correct news headlines?

### *Society Alarmed by Misinformation*

Not only exposure to actual misinformation can distort people's credibility ratings, potentially also the heightened salience of misinformation as an imminent threat might trigger a deception bias, obstructing the default truth bias. Despite that misinformation only plays a minor role in audiences' average media diets (Acerbi, Altay, and Mercier 2022), it is a prominent topic of public discourse (Van Duyn and Collier 2019). News media have constantly covered the topic of misinformation, not only by fact-checking news items but also by talking about it as a prominent threat to society. Reasons for this constant coverage may be media's focus on issues characterized as exceptional—i.e., distortion bias (Entman 2007)—and negative—i.e., negativity bias (Soroka and McAdams 2015). Misinformation's rarity and potential negative impact make it a highly newsworthy issue, explaining why the threat of false information might gain disproportionate media attention. Research, for example, shows how growing up around public debates about “fake news” can lead to differences in media trust evaluations among generations (Brosius, Ohme, and de Vreese 2021) and how exposure to elite discourse about the prevalence of misinformation leads to lower levels of media trust and less accurate identification of real news (Van Duyn and Collier 2019). If media coverage consistently primes people to misinformation being a prominent threat to society, a deception bias may be the consequence.

Extant research has been concerned with the puzzle of to what extent attempts to educate people to critically navigating their news environment generate healthy skepticism towards information or cause harmful cynicism (Hobbs and McGee 2014; Mihailidis 2009; Mihailidis and Thevenin 2013; Mihailidis and Viotty 2017). It has increasingly become apparent that media analysis skills extend beyond the ability to read messages, including the need to situate information in relation to broader social, cultural, and political contexts (Mihailidis and Viotty 2017). If the ability to effectively use media to exercise democratic rights is ignored in educating critical processing of media, interventions and literacy programs run the risk of fostering cynical dispositions instead of a more nuanced ecological understanding of media (Mihailidis and Thevenin 2013). Rather than stimulating critical reflection of why media work as they do, it may breed an unhealthy defensive view. Thus, at times (Hobbs et al. 2013; Hobbs and McGee 2014), those who are thought or reminded to critically process media messages might simultaneously adopt a cynical attitude toward media credibility and relevance (Mihailidis 2009; Mihailidis and Thevenin 2013). Especially when warnings are framed as tricks designed to mislead people, they may contribute to higher levels of generalized media cynicism (Hobbs and McGee 2014).

Caution is especially warranted in today's era of eroding public trust and growing cynicism in light of the drastic changes—i.e., distributed propaganda, hijacking of local news, and reifying polarization—that media institutions have undergone in the past decades (Mihailidis and Foster 2021). This reflects challenges for educating audiences to be more critical: creating skeptical news consumers without fostering undue cynicism that may hinder generalized distrust in news content (Tully, Vraga, and Smithson 2020). Critical thinking should therefore be accompanied by an awareness of the necessity of media's

democratic role, otherwise critical thought can easily become cynical thought (Mihailidis and Thevenin 2013).

This study aims to translate such unintended effects, regarding potentially cultivating cynical media views while attempting to stimulate critical information processing, to the context of misinformation. We therefore explore the potential *negative spillover effect* of attempts to fight misinformation. Here we understand spillover effects as an unintended ripple effect of efforts to combat misinformation on other domains, namely people's general evaluation of the news and their trust in factually accurate information. Fighting misinformation might have an indirect negative effect on perceptions of general news credibility as a consequence of well-intentioned corrections of or warnings for misinformation that, as a side product, make the threat of misinformation as a risk to our information ecosystem more salient. This salience mainly persists through certain communicative interventions or warnings. We thus ask if communicative attempts to fight misinformation can spill over to ratings of truthful information, where people are primed towards the risk of being misinformed which may distort their news credibility assessment. Especially since priming effects are found to transfer to relevant concepts (Petty and Jarvis 1996), being primed with the threat of being misled might transfer to the assessment of reliable news. Thus, although the effect of misinformation as an information genre is debatable, it may have more wide-ranging consequences through its prevalence in public and media discourse (Van Duyn and Collier 2019).

Against this backdrop, we examine if interventions alerting to the presence or dangers of false information—i.e., like factchecks or general warnings—might have a negative spillover effect as they prime misinformation as a prominent threat and trigger a deception bias if it comes to judgement of real news. Contrary to the truth-default theory, citizens might grow overly skeptical toward news credibility in the context of a public debate that generally understands, and potentially overestimates, misinformation as a major challenge. We ask:

**RQ2:** Can (a) misinformation corrections and (b) warnings for the omnipresence of misinformation lower credibility ratings of factually correct news headlines?

### ***News Media Literacy Interventions as a Solution?***

Compared to corrections of or warnings for misinformation, news media literacy (NML) interventions can be used as educative efforts to increase trust in reliable information and improving the quality of the news ecosystem more generally (Acerbi, Altay, and Mercier 2022). NML literature is largely concerned with how communicative interventions can improve citizens' skills and knowledge regarding the process of how news is produced and how to navigate the information environment in a thoughtful manner (Ashley, Maks, and Craft 2017; Jones-Jang, Mortensen, and Liu 2021). An essential part of media literacy relates to audiences' understanding of biases in news production and consumption, a skill crucial for deception detection (Flynn, Nyhan, and Reifler 2017). Communicative efforts can, for example, help news consumers reflect on their own biases and those in news production when processing information (van der Meer and Hameleers 2022; Vraga et al. 2021) and have been proven to stimulate desired news selection behavior (van der Meer and Hameleers 2020; Vraga and Tully 2019). In the context of

misinformation, it has been argued that literacy education would better equip audiences with resilience to information disorder, more so than reactive resources like fact-checking and verification tools (Mcdougall 2019).

Extant research concludes that improving trust in reliable sources is a challenging task but also showcases how NML interventions can, under certain circumstances, help to stimulate the acceptance of reliable news. Relying on a simulation study, Acerbi, Altay, and Mercier (2022) opt for devoting more efforts to improving acceptance of reliable news instead of fighting misinformation. They show that small increases in accepting real news improve the information score more than decreasing the acceptance of misinformation. Moreover, they argue that the effectiveness of interventions directed at fighting misinformation are destined to be minuscule in the greater context, especially compared to the effects of stimulating greater acceptance of reliable information (also see Guess et al. 2020). Moreover, transparency boxes are found to increase credibility perceptions (Masullo et al. 2021), and applying fact-checking techniques showed effective when deciding whether information was scientifically valid or not (Panizza et al. 2022). Moreover, digital NML interventions have potential in terms of decreasing perceived accuracy of false information and help audiences better distinguish it from factual news (e.g., Guess et al. 2020; Hameleers 2020). Next, Jungherr and Rauchfleisch (2022) test whether balanced warnings, which emphasize the presence of disinformation while accounting for that its dangers can be overestimated, offer an alternative to indiscriminative misinformation warnings. They show how such balanced accounts can lower the perceived threat level of misinformation.

Yet, NML interventions targeted at misinformation are not always successful. For example, providing audiences with baseline statistics about the level of misinformation consumption was found to rather exacerbate than downplay perceived prevalence and influence of misinformation (Lyons, Merola, and Reifler 2020). Exposure to digital short-form NML interventions were also found to fail enhancing effectiveness of misinformation correction and rather generated cynicism (Vraga, Tully, and Bode 2022) and a NML video discussing how credible news and false information are often mixed up online did not inoculate people to false health information (Vraga, Bode, and Tully 2022). In addition, exposure to interventions to fight misinformation can reduce the perceived accuracy of not only false headlines but also mainstream news headlines (Guess et al. 2020). Hence, the value of NML interventions is predicated in its ability to both reduce the harms of misinformation without spilling over to news credibility.

To translate the above findings to the context of credibility assessment of factual information, this study explores whether NML interventions can be successful in avoiding unintended spillover effects of fighting misinformation. Instead of training citizens to constantly verify the veracity of content they come across, it might be more fruitful to encourage them to trust reliable sources (Altay 2022). Interventions that reconcile trust in accurate information, by offering contextual information about the relative threats of misinformation, might help to marginalize fear for deception and the relativization of the omnipresence of misinformation might consequentially improve accuracy of people's news credibility assessment. Yet, as discussed above, such interventions might be complex at best, just like fact-checks and warnings for misinformation, NML interventions can be unsuccessful or even backfire as these interventions can unintentionally bias



audience perception rather than generate healthy skepticism. Therefore, we ask the open question:

**RQ3:** Can a NML intervention that relativized the threats of misinformation in the context of the dominance of accurate information increase people's credibility rating of factually correct news headlines?

## Method

In an effort to understand if fighting misinformation can have negative spillover effects, an experimental design was set up. The study is designed to understand how single interventions cause an informational, direct effect wherein exposure primes certain way of information assessment. The general approach of this study is to compare the veracity rating of factually correct news headlines after exposure to different stimuli. The approach followed a 5 (exposure to: misinformation article, factcheck after misinformation article, a NML intervention aimed to increase credibility accesses of real news, or a control message) X 2 (issue: climate change misinformation or general misinformation regarding elites) between-subjects design (see Appendix A for the experimental flow chart). The two issues included in the design tend to be contested and therefore often subject to misinformation. Climate change is subject to partisan interpretations and polarization, which also makes it vulnerable to de-legitimizing attacks. The issue blaming elite actors for misinformation is in line with the omnipresence of misinformation accusations and fake news labels as a genre and label of misinformation at the same time (Egelhofer and Lecheler 2019). In that sense, we focus on two issues that are most likely to be surrounded by de-legitimizing discourse. The first issue has a specific focus on climate change for both the stimuli and news headlines while the second design takes a more general approach to test the spillover effect of (fighting) misinformation on decreasing the credibility of news for different topics.

We focus on the US as this setting is regarded as having a lower resilience to disinformation due to high levels of polarization and declining levels of trust (Humprecht et al. 2022). It is thus a "most likely" context for the problems related to the deception and truth biases studied in this paper. In the US media environment, it used to be well-established that most news is roughly accurate, though possibly slanted. Therefore, the US provide a valid case to study if moral panic over misinformation can indeed spillover to other domains. Yet, it needs to be noted that this is not the case everywhere. For example, the Nordic countries' high institutional trust, which includes the media, seems to more effectively immunize against misinformation panics; whereas in several countries in the post-Soviet space, concern about inaccurate news is a more long-established pattern that should readily absorb misinformation moral panics as nothing new.

## Sample

US respondents were recruited online in January 2022 via sampling company Dynata. A total of 1,305 fully completed the experiment and were randomly assigned to the two topics, resulting in 654 for the issue of climate change and 651 for the elite issue. Quotas were in place for age and gender crossed, and education. On average, participants



were 46.71 years old ( $SD = 20.64$ ) and 52.34% identified as female. In total, 14.79% were lower educated, 51.42% had a moderate education level, 33.79% were higher educated. 586 of them identified as Democrats, 446 as Republicans, and the remaining 273 as Independents.

## Procedure

The study and all principles reported here received ethical approval from the institutional review board. After the introduction and consent form, participants completed the pre-treatment questions for the moderator and control variables. On the next page, they were introduced to the stimuli. For the NML intervention condition, respondents were informed that they would read a public service announcement (PSA) on the selection of news from the Media Literacy Coalition, issued that week (Vraga and Tully 2019). In the other conditions, the general instruction asked participants to carefully read the next article that was published online that week. All stimuli were based on existing articles on topics of fake news or misinformation corrections/warnings.

The stimuli for the topic of climate change looked as follows: First, in the misinformation exposure condition, respondents read an article that falsely denies that human interference causes climate change (see Appendix B). The misinformation stimuli contained statements like: “Contrary to what many argue, human activity is not a significant contributing factor in changing global temperatures.” Second, the misinformation warning condition showed an article that argues how misinformation on climate change is omnipresent these days (see Appendix C), including sentences like: “our current information environment and understanding of climate change is threatened by the spread of misinformation.” Third, in the correction condition, respondents were first exposed to the same stimuli arguing that human interference do not cause climate change as in the misinformation condition followed by a fact-check effort stating that the majority of claims made in the article are incorrect (see Appendix D)—e.g., “Fact-checking efforts reveal that claims stating that climate change is not caused by human interference are completely false.” We consider this condition a less direct manipulation, compared to the second experimental group, since respondents are implicitly reminded of the threat of misinformation by exposure to false and corrective information. Fourth, in the final experimental condition, respondents read an NML intervention arguing that, despite its potential danger, the volume of misinformation is relatively low and should not by default lower trust in all news outlets and content (see Appendix E)—e.g., “Yet, despite the fact that misinformation can have severe consequences, the percentage of people being exposed to false or misleading information is relatively small.” Finally, in the control group, respondents read about a historical painting being restored (see Appendix F). All articles were comparable in number of words and respondents had to, at least, spend 30 s on the page for sufficient reading time to process the stimuli.

For the second topic, we look at a more overarching misinformation context that could spill over to all types of news, namely how the media are run by a small group of elites. For this issue, the misinformation condition stated how corrupt elite control the media system—e.g., “The media industry is created based on a propaganda model, where news is controlled by corporate-owned news media that produce news content to achieve a profit by convincing people to think in a certain way that satisfies the common interests of the

powerful elites.” (see Appendix G). While a large proportion of our mainstream media is indeed owned by large corporations, this message is labeled as misinformation and an attempt to mislead as it refers to how a small group aims to manipulate and control the public via spreading goal-directed information using the mainstream media they own. This manipulation is in line with the delegitimizing narrative typically used in disinformation accusations or “fake news” labels (Egelhofer and Lecheler 2019). As the accusation targeting powerful media elites allegedly controlling public opinion and deceiving the public for their own profit is not based on expert knowledge and/or factually accurate information (Vraga and Bode 2020), we consider it as misinformation. The message actively blames the media elite for “manipulation” and “the creation of narratives used to control public opinion.” Taking into consideration that delegitimizing disinformation labels that blame mainstream media for deception are often not driven by facts, expert knowledge, or rational arguments, but rather by a fact-free anti-establishment logic, they can be considered as mis- or disinformation in their own right (Hameleers 2023). That being said, just like many instances of misinformation, they may contain some factually accurate basis on which the deception is founded. The misinformation warning entailed a more general statement that warned of the fast spread of false information in our information environment (see Appendix H)—e.g., “our current information environment is threatened by the spread of misinformation.” In the correction condition, respondents again saw a factcheck stating that the majority of claims of the misinformation article (same as shown in the misinformation condition) they saw before were incorrect (see Appendix I)—e.g., “There is no proof for the allegation that mainstream media, corporations or political actors deliberately fabricate information for political or financial gains.” The NML intervention and control condition were the same as for the other topic.

After exposure to one of the experimental stimuli or control group, participants were instructed that, on the next pages, they would read four news headlines with their lead that recently circulated online. Since we aim to test if attempts to fight misinformation have a negative spillover effect on general perceptions of news credibility, only factual headlines were included. Thus, we are interested in assessing whether people accurately perceived factual news to be credible (even when the threat of misinformation has been made salient). Providing respondents with only factual headlines better reflect people’s real-world information environment (Acerbi, Altay, and Mercier 2022) than, for example, showing the same amount of factual and false headlines. Moreover, exposing respondents to fake information might be too much of a prime that other news headlines could be inaccurate. For each article, respondents were asked to rate on a 0–100 scale how credible they thought the information in the news item was. We refrained from asking whether headlines were considered real or fake, since such a primed question can potentially, as argued by truth-default theory, trigger suspicion of deception. Yet, it needs to be acknowledged that asking respondents to rate information credibility might still be a potential trigger to raises suspicions. All news items (source information omitted) were based on real news headlines from outlets like CNN, NBC News, and Fox News and were selected based on their factual accuracy. For the topic of climate change, all four headlines related to climate change and somehow were in line with the position that human interference causes climate change, which is contrary to the misinformation item (see Appendix J). For the topic of elites controlling the media, different types of news items were presented (see Appendix K) because the rating for the climate

change topic might be prone to demand effects since the stimuli and headlines are both on the topic of climate change. If the stimuli argue that climate change is not induced by human behavior or that climate change misinformation is everywhere, respondents' ratings of the headlines might be too focused on ensuring their ratings are in line with the article they were just exposed to. Therefore, the headlines of the second topic contained different topics: climate change, politics, science, and health.

To finalize the questionnaire, participants responded to the demographic questions, manipulation check, and were elaborately debriefed. They returned to the environment of the research company collecting the data after having read the debriefing that debunked all misinformed claims made in the experiment.

## Measures

*Credibility rating.* To obtain participants' accuracy score, their credibility ratings (0-100 scale) of all four news items were averaged ( $M = 55.73$ ;  $SD = 22.54$ ). Since all news items contained factual information, a higher score indicates more accurate credibility assessment, allowing to test if fear for misinformation has unintended consequences for evaluating reliable news. By asking respondents to rate different news articles for their credibility to assess respondents' accurate judgment, we followed prior work for calculating detection/credibility accuracy (e.g., Levine, Shaw, and Shulman 2010; Luo, Hancock, and Markowitz 2022).

*Control variables.* Several control variables were included, all measured on a 7-point Likert scale. Overall media trust was measured with one item asking whether people trust the news media most of the time ( $M = 3.98$ ;  $SD = 1.81$ ). Prior misinformation beliefs were measured with four items ( $M = 3.92$ ;  $SD = 1.64$ ; Cronbach's  $\alpha = .76$ ), depending on which of the two issues respondents were exposed to, either asked about their agreement with that climate change is not caused by human activities (e.g., "There is no scientific consensus that global warming is caused by humans" and "Human activity is primarily causing climate change (i.e., global warming)") or that powerful elites control news outlets (e.g., "Powerful elites manipulate media content for their own profit" and "Politicians and media try to control the public debate to conceal real problems"). Political ideology was measured at the end of the questionnaire with the demographic items, asking whether they think of themselves as a Republican, a Democrat, or an Independent. Hostile media perception (HMP) was measured with three items on a 7-point Likert scale ( $M = 4.24$ ;  $SD = 1.56$ ; Cronbach's  $\alpha = .96$ ), asking about whether they perceived that US news media are generally biased against their views on issues (Hwang, Pan, and Sun 2008).

*Manipulation check.* The manipulation-check question was one item with five different answer categories, each category representing one of the conditions. In total, 78% answered correctly. The question asked what type of article they read e.g., one argued that climate change is not caused by human activities or message from Media Literacy Coalition that we should not lose trust in all news because of fear for misinformation. Chi-Square test indicated that the manipulation was successful ( $\chi^2(16) = 2187.7$ ,  $p < 0.001$ ). Besides the manipulation check, an open-ended question, where respondents reflected on the stimulus they were exposed to, served as a qualitative check for message clarity and unambiguity.

## Results

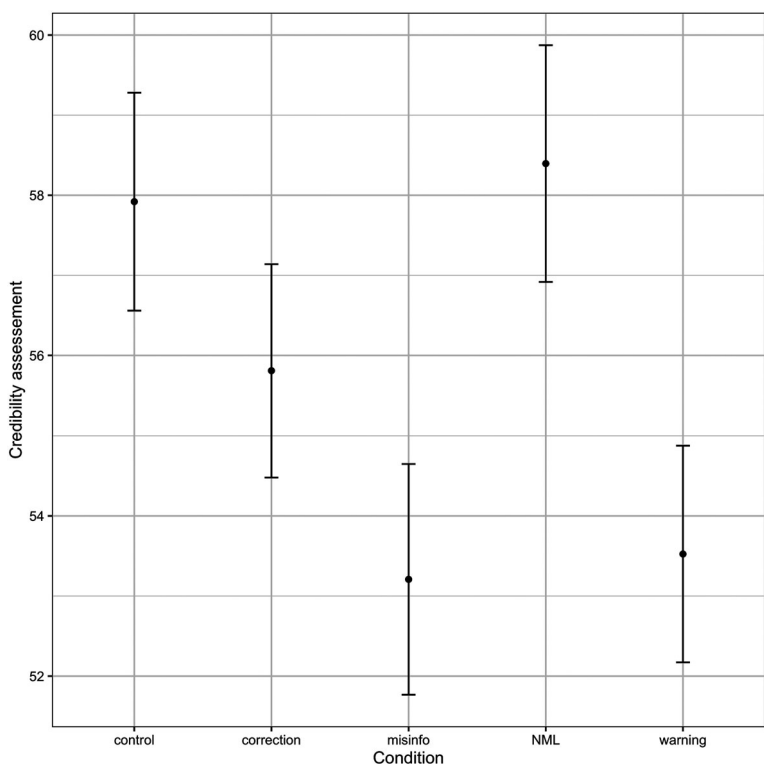
To test whether misinformation exposure decreases credibility rating of true news (RQ1), regression models were run with stimuli exposure as independent variable and credibility scores as dependent variable. Table 1, model 1, shows how people exposed to misinformation rated true news items as significantly less credible than those in the control group. This indicates that participants exposed to misinformation rated subsequent true news as less credible. This difference still holds if we control for hostile media perception, overall media trust, misinformation beliefs, and political ideology (see, Table 1, Model 2). Appendix L details how these differences hold across both topics—i.e., climate change and elites controlling the media—when we run the analyses separately (Models 2-3) and that no moderation effect of topics was found (Model 1). Yet, the effect of the misinformation conditions seems to be primarily driven by the climate change condition. No interaction between misinformation exposure and prior misinformation beliefs was observed, suggesting that misinformation can lower credibility rating for both those who are more likely to believe the misinformation and those who are not. The difference in credibility judgement across the different conditions are also visualized in Figure 1.

RQ2 asked about whether attempts to fight misinformation can affect people's credibility ratings of true information. Table 1 shows how participants in the condition of the fact check (RQ2a) did not rate the credibility of the news items significantly different compared to those in the control group. There was thus no unintended spill-over effect of the fact check on the rating of factually accurate news. Yet, when exposed to the warning for

**Table 1.** Regression models explaining news credibility assessment.

	<i>Dependent variable:</i>	
	News credibility assessment	
	(1)	(2)
Stimuli: Misinformation <sup>a</sup>	−4.87* (1.96)	−4.28* (1.69)
Stimuli: Correction <sup>a</sup>	−2.26 (1.95)	−2.93 <sup>†</sup> (1.68)
Stimuli: Warning <sup>a</sup>	−4.60* (1.95)	−4.70** (1.68)
Stimuli: NML intervention <sup>a</sup>	0.48 (1.99)	−0.44 (1.71)
Hostile media perception (HMP)		3.08*** (0.40)
Media trust		3.96*** (0.33)
Misinformation beliefs		−4.24*** (0.49)
Political ideology <sup>b</sup>		1.87*** (0.31)
Topic (Elite) <sup>c</sup>	−4.95*** (1.24)	2.58* (1.33)
Constant	65.44*** (2.34)	34.23*** (3.22)
Observations	1305	1304
R <sup>2</sup>	0.02	0.28
Adjusted R <sup>2</sup>	0.02	0.27
Residual Std. Error	22.35 (df = 1299)	19.24 (df = 1294)
F Statistic	5.58*** (df = 5; 1299)	55.15*** (df = 9; 1294)

Note: <sup>†</sup> $p < 0.1$ ; \* $p < 0.05$ ; \*\* $p < 0.01$ ; <sup>a</sup> Reference group is control condition; <sup>b</sup> Ranging from Republican (low score) to Democrat (high score); <sup>c</sup> Reference group is the other topic: Climate change.



**Figure 1.** Mean difference in credibility assessment across conditions.

misinformation (RQ2b), respondents rated the news headlines as less credible, which also means that their credibility assessment was less accurate as a consequence of this warning.

RQ3 asked whether exposure to NML interventions can increase the accuracy of credibility assessments in a context of fear of misinformation. Table 1 shows how exposure to a NML intervention did not change credibility ratings compared to the control group. However, when misinformation (Table 2, Model 1) or warning for misinformation (Table 2, Model 3) are set as a reference group, it was observed that those in the NML intervention condition had a more accurate news assessment. These findings suggest that an NML intervention does not per se increase the accuracy of news assessment but can be important in a context where audiences are frequently exposed to misinformation or warned for its omnipresence.

Discussion

This study explored whether attempts to fight misinformation can have a negative spillover effect by priming disproportionate levels of suspicion. Efforts like misinformation warnings might prime general deception in information, despite most news being reliable and based on factual information. Misinformation may not only be harmful to democracy as an informational crisis, but its frequent use as a label or prominence in public and media discourse may have negative consequences for credibility assessment of authentic

**Table 2.** Regression models predicting news credibility assessment.

	<i>Dependent variable:</i>		
	news credibility assessment		
	(1)	(2)	(3)
Stimuli: Control	4.28* (1.69)	2.93 <sup>†</sup> (1.68)	4.70** (1.68)
Stimuli: misinformation	RG	−1.36	0.42
Stimuli: Correction	1.36 (1.67)	RG	1.78 (1.66)
Stimuli: Warning	−0.42 (1.68)	−1.78 (1.66)	RG
Stimuli: NML intervention	3.85** (1.71)	2.49 (1.69)	4.27* (1.70)
Hostile media perception	3.08*** (0.40)	3.08*** (0.40)	3.08*** (0.40)
Media trust	3.96*** (0.33)	3.96*** (0.33)	3.96*** (0.33)
Misinformation beliefs	−4.24*** (0.49)	−4.24*** (0.49)	−4.24*** (0.49)
Political ideology <sup>a</sup>	1.87*** (0.31)	1.87*** (0.31)	1.87*** (0.31)
Topic (Elite) <sup>b</sup>	2.58 <sup>†</sup> (1.33)	2.58* (1.33)	2.58 <sup>†</sup> (1.33)
Constant	29.94*** (3.20)	31.30*** (3.24)	29.53*** (3.24)
Observations	1,304	1,304	1,304
R <sup>2</sup>	0.28	0.28	0.28
Adjusted R <sup>2</sup>	0.27	0.27	0.27
Residual Std. Error (df = 1294)	19.24	19.24	19.24
F Statistic (df = 9; 1294)	55.15***	55.15****	55.15****

Note: <sup>†</sup> $p < 0.1$ ; \* $p < 0.05$ ; \*\* $p < 0.01$ ; RG: *reference group*; <sup>a</sup> Ranging from Republican (low score) to Democrat (high score);

<sup>b</sup> Reference group is the other topic: Climate change.

information. Against this backdrop, we used an online survey experiment to explore how misinformation exposure, pre-bunking forewarnings, debunking fact-checks, and NML interventions relativizing the misinformation threat affected people's ratings of factually correct information.

Our key findings show that exposure to both misinformation and a warning for misinformation negatively impact the credibility of factually accurate information. Thus, we confirmed how exposure to actual misinformation can cause misperceptions, in the sense that it lowers accurate credibility assessment. Yet, also talking about misinformation—i.e., being alerted about misinformation as a threat—may have a negative spill-over effect on how news users rate truthful information in a post-factual information era. Such public debates about the omnipresence of misinformation, hence, have the potential to prime a general context where audiences have become disproportionately distrustful of all information. These findings resonate with previous research that concluded how elite discourse about fake news can lower media trust and accuracy of identifying real news (Van Duyn and Collier 2019). On the contrary, we did not observe that exposure to misinformation that was later debunked by a fact check decreased credibility perception of real news. Potentially the presence of a fact check makes the threat of misinformation less prominent and therefore does not prime a deception default mindset.

As a theoretical starting point, this study asked whether a truth-default bias would hold for news assessment in a context of misinformation. Unlike the robust observation for a

truth-bias in interpersonal communication (Levine 2014), our study showcased how respondents are less likely to believe that factual information is credible in light of exposure to misinformation or misinformation warnings. Seemingly, even misinformation warnings can serve as a trigger where communication is by default judged as more likely to be false, creating a deception-bias instead of a truth-bias. The truth-default theory may therefore not be a valid and applicable theory to understand news credibility assessment in a fragmented high-choice information era, where news, misinformation and pseudo-information can come from a plethora of alternative sources. In addition, given that mis- and disinformation mimicking the news values of actual information and are presented alongside accurate content, while real news has frequently been accused of being fake news (Egelhofer and Lecheler 2019; Waisbord 2018), it is very challenging for citizens to know whom to trust. The deception-bias we observed may therefore reflect the presumption of news being generally difficult to trust. This finding corresponds with global longitudinal data showing that media trust is indeed declining in most countries.

In a next step, we provided a first exploration of whether NML interventions can counter a potential deception-bias in news assessment. The findings indicated that the effectiveness of a PSA that relativizes the threat of misinformation is not so straightforward. Compared to the control condition, the NML intervention did not improve the accuracy of people's credibility rating. Yet, in comparison with exposure to misinformation or forewarning, the intervention helped respondents better assess the credibility of news. This might suggest that in a news environment where misinformation is present or in a context of discourse about the threat of misinformation, NML interventions about the salience of misinformation can help to increase accurate assessment of real news. Overall, and in line with previous research (Tully, Vraga, and Bode 2020; van der Meer and Hamel-eers 2020; Vraga, Bode, and Tully 2020), our findings highlight the complexity of stimulating media literacy with communicative interventions. Additional work is required, for example, on testing whether repeated exposure would be more successful or to what extent content and strategy of the specific PSAs determines its effectiveness. Nevertheless, the conditional effectiveness might be reason for some optimism, especially since interventions aimed at increasing trust in reliable news sources were predicted to primarily improve the quality of the information environment rather than those efforts set out to reduce the acceptance or spread of false information (Acerbi, Altay, and Mercier 2022).

It is important to reflect on the overall credibility assessment ratings of true headlines observed in this study. The average credibility score, also for participants exposed to the control condition, was only mildly positive and just above the mid-point of the scale. These scores are in line with previous studies (e.g., Luo, Hancock, and Markowitz 2022). More specifically, a meta-analysis of interpersonal deception studies revealed 54% accuracy in detecting deception with equal lie-truth base-rates, here people more accurately judge true information as honest (61%) than lies as deceptive (47%; Bond and DePaulo 2006). These findings can be concerning as they indicate poor credibility assessment of factual news. The main conclusion to be drawn here is that the accurate credibility assessment of news is at best a highly complex task (McGrew et al. 2018; Pennycook and Rand 2020), especially in the contemporary media environment public trust in information is eroding (Mihailidis and Foster 2021) and people become disproportionally prone to believe that information can be blatantly false. Such general suspiciousness might



coincide with high estimations of misinformation salience on social media, driving a deception bias in online-dominated news landscape. Since fake news exposure is in decline on social network sites like Facebook (Allcott, Gentzkow, and Yu 2019), people's deception biases might grow even more problematic since their skepticism becomes even less warranted. Therefore, it is crucial for media literacy education to teach people to clearly understand baseline statistics of misinformation salience without priming a deception default attitude.

Theoretically, the findings of our paper suggest that misinformation research should not be limited to studying the effects of misinformation as genre or label (e.g., Egelhofer and Lecheler 2019), but partly shift to a focus on the information environment surrounding misinformation (e.g., Brosius, Ohme, and de Vreese 2021)—including the unintended side-effects of remedies. Our finding that the truth-default of credibility assessment may not be valid in today's information environment implies that we should carefully consider how real news and accurate information are processed alongside disinformation, warning labels, and fake news beliefs and accusations. Hence, the impact of misinformation should not be approached as the individual effects of individual messages on responses to these messages, but as a wider "disorder" that may also impact the ways in which people assess truthful information (e.g., Bennett and Livingston 2018).

Practically, our findings imply that de- or pre-bunking initiatives that aim to prevent or counter misinformation should do more than simply warn people of an incoming attack. Although such interventions have a proven impact on reducing misperceptions or instilling resistance to misinformation (e.g., Hameleers 2020; Roozenbeek and van der Linden 2019; Vraga et al. 2021), they may have unintended effects on perceptions of real information by priming (disproportionate) suspicion. The effectiveness of interventions should not only be considered in terms of weakening the impact of misinformation, but also as bolstering or at least consolidating the perceived accuracy and trustworthiness of factually correct information. Therefore, we suggest paying specific attention to restoring trust in facts, expert sources, and verified evidence in efforts to fight misinformation. In addition, the threat of misinformation should be relativized by showing news users that most information may be trustworthy, also offering them efficacy beliefs by revealing where trustworthy information can be obtained. We are not arguing to dismantle initiatives to counter misinformation effects, yet overall more efforts should be devoted to increase trust in reliable news sources (Acerbi, Altay, and Mercier 2022). Concretely, NML interventions may include statements on the relative salience of fake versus real news and offer concrete suggestions on what type of information is most likely to be trustworthy and accurate. On a more practical level, a dynamic score for the level of identified misinformation in a specific news environment may help making people aware of the possibility but still helps to realistically assess chances to encounter a factually incorrect news item.

The current study is not without limitations. First, the artificial news exposure in our experimental set up does not capture the complexity of people's individual news environment. Respondents were forcefully exposed to, for example, misinformation or warnings. This does not accurately reflect the extent to which one's digital news environment is polluted with misinformation or characterized by prominent discourse of the threat of misinformation. Future research may need to take selective exposure to misinformation (warnings) into account and predict which segments of the news audience are most likely to avoid or select remedies as well as the disease. Likewise, under real-world

conditions, source cues are omnipresent and a primary credibility cue. In the current design, these source cues were left out to study the effects of interventions in isolation, but this might have affected the external validity of the study. Second, the study was only limited to a certain number of stimuli. For example, only one type of NML intervention was used. For future research, it would be worthwhile to explore the effect of different types of interventions. For example, previous research showed how using either descriptive or normative language in NML interventions differently affected people with other ideological beliefs, while tailored interventions were successful across the board (van der Meer and Hameleers 2020) or how a combination of or balance between different forms of warning can have the desired effect (Jungherr and Rauchfleisch 2022). Moreover, we only included two topics of misinformation. These stimuli also differed on several levels which future research should test for their effect on credibility assessment separately. For example, the misinformation stimuli across the topics differed in level of inaccuracy, where denying human interference might be more factually incorrect than accusations of elites owning and manipulating mainstream media. Additionally, the elite misinformation condition included a delegitimizing misinformation label. Although this does follow the delegitimizing and antagonist narrative of many disinformation campaigns, it includes a misinformation accusation with information that is not based on expert knowledge and empirical evidence. Related, we exclusively let participants rate factually accurate headlines. Although this is closer to people's real-world experience and in line with our theoretical interest to map the unintended spillover effect of misinformation (discourses) on the trustworthiness of real news, future research may use a mixture of false and true headlines—whilst taking into account that factually accurate information is much more prevalent in people's news diets. Finally, it is a challenging task to contextualize and fully understand the relative harm of spillover effects. It is difficult to assess whether misperceptions caused by exposure to false information are more consequential in societies than decreased credibility assessment of legitimate news in light of fear of misinformation. Both forms of impact are related to different potential outcomes (persistence of misperceptions versus low media trust) that can both impact democracy differently. Nevertheless, the study provides important insights into how the threats posed by the misinformation order may not just be remedied by fighting fake information, but also by reestablishing trust in legitimate news institutions.

## Funding Details

This work was supported by a VENI grant from the Netherlands Organization for Scientific Research (NWO) [grant number 016.Veni.195.067].

## Disclosure Statement

No potential conflict of interest was reported by the author(s).

## References

- Acerbi, A., S. Altay, and H. Mercier. 2022. "Research Note: Fighting Misinformation or Fighting for Information?" *Harvard Kennedy School Misinformation Review*. doi:[10.37016/mr-2020-87](https://doi.org/10.37016/mr-2020-87).

- Allcott, H., M. Gentzkow, and C. Yu. 2019. "Trends in the Diffusion of Misinformation on Social Media." *Research & Politics* 6 (2): 205316801984855. doi:[10.1177/2053168019848554](https://doi.org/10.1177/2053168019848554).
- Allen, J., B. Howland, M. Mobius, D. Rothschild, and D. J. Watts. 2020. "Evaluating the Fake News Problem at the Scale of the Information Ecosystem." *Science Advances* 6 (14): eaay3539. doi:[10.1126/sciadv.aay3539](https://doi.org/10.1126/sciadv.aay3539).
- Altay, S. 2022. "How Effective Are Interventions Against Misinformation? [Preprint]." *PsyArXiv*, doi:[10.31234/osf.io/sm3vk](https://doi.org/10.31234/osf.io/sm3vk).
- Altay, S., E. de Araujo, and H. Mercier. 2022. "'If This Account is True, It is Most Enormously Wonderful': Interestingness-If-True and the Sharing of True and False News." *Digital Journalism* 10 (3): 373–394. doi:[10.1080/21670811.2021.1941163](https://doi.org/10.1080/21670811.2021.1941163).
- Appelman, A., and S. S. Sundar. 2016. "Measuring Message Credibility: Construction and Validation of an Exclusive Scale." *Journalism & Mass Communication Quarterly* 93 (1): 59–79. doi:[10.1177/1077699015606057](https://doi.org/10.1177/1077699015606057).
- Ashley, S., A. Maksl, and S. Craft. 2017. "News Media Literacy and Political Engagement: What's the Connection?." *Journal of Media Literacy Education* 9 (1): 79–98.
- Bennett, W. L., and S. Livingston. 2018. "The Disinformation Order: Disruptive Communication and the Decline of Democratic Institutions." *European Journal of Communication* 33 (2): 122–139. doi:[10.1177/0267323118760317](https://doi.org/10.1177/0267323118760317).
- Bond, C. F., and B. M. DePaulo. 2006. "Accuracy of Deception Judgments." *Personality and Social Psychology Review* 10 (3): 214–234. doi:[10.1207/s15327957pspr1003\\_2](https://doi.org/10.1207/s15327957pspr1003_2).
- Bond, G. D., D. M. Malloy, E. A. Arias, S. N. Nunn, and L. A. Thompson. 2005. "Lie-Biased Decision Making in Prison." *Communication Reports* 18 (1–2): 9–19. doi:[10.1080/08934210500084180](https://doi.org/10.1080/08934210500084180).
- Brosius, A., J. Ohme, and C. H. de Vreese. 2021. "Generational Gaps in Media Trust and its Antecedents in Europe." *The International Journal of Press/Politics* 194016122110394. doi:[10.1177/19401612211039440](https://doi.org/10.1177/19401612211039440).
- Buller, D. B., and J. K. Burgoon. 1996. "Interpersonal Deception Theory." *Communication Theory* 6 (3): 203–242. doi:[10.1111/j.1468-2885.1996.tb00127.x](https://doi.org/10.1111/j.1468-2885.1996.tb00127.x).
- Burgoon, J. K. 2015. "Rejoinder to Levine, Clare et al.'s Comparison of the Park-Levine Probability Model Versus Interpersonal Deception Theory: Application to Deception Detection: Rejoinder to Levine, Clare, Green, Serota, & Park, 2014." *Human Communication Research* 41 (3): 327–349. doi:[10.1111/hcre.12065](https://doi.org/10.1111/hcre.12065).
- Clare, D. D., and T. R. Levine. 2019. "Documenting the Truth-Default: The Low Frequency of Spontaneous Unprompted Veracity Assessments in Deception Detection." *Human Communication Research* 45 (3): 286–308. doi:[10.1093/hcr/hqz001](https://doi.org/10.1093/hcr/hqz001).
- Egelhofer, J. L., and S. Lecheler. 2019. "Fake News as a Two-Dimensional Phenomenon: A Framework and Research Agenda." *Annals of the International Communication Association* 43 (2): 97–116. doi:[10.1080/23808985.2019.1602782](https://doi.org/10.1080/23808985.2019.1602782).
- Entman, R. M. 2007. "Framing Bias: Media in the Distribution of Power." *Journal of Communication* 57 (1): 163–173. doi:[10.1111/j.1460-2466.2006.00336.x](https://doi.org/10.1111/j.1460-2466.2006.00336.x).
- Flynn, D. J., B. Nyhan, and J. Reifler. 2017. "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics: Nature and Origins of Misperceptions." *Political Psychology* 38: 127–150. doi:[10.1111/pops.12394](https://doi.org/10.1111/pops.12394).
- Gilbert, D. T. 1991. "How Mental Systems Believe." *American Psychologist* 46 (2): 107–119. doi:[10.1037/0003-066X.46.2.107](https://doi.org/10.1037/0003-066X.46.2.107).
- Grinberg, N., K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer. 2019. "Fake News on Twitter During the 2016 U." *S. Presidential Election. Science* 363 (6425): 374–378. doi:[10.1126/science.aau2706](https://doi.org/10.1126/science.aau2706).
- Guess, A. M., M. Lerner, B. Lyons, J. M. Montgomery, B. Nyhan, J. Reifler, and N. Sircar. 2020. "A Digital Media Literacy Intervention Increases Discernment Between Mainstream and False News in the United States and India." *Proceedings of the National Academy of Sciences* 117 (27): 15536–15545. doi:[10.1073/pnas.1920498117](https://doi.org/10.1073/pnas.1920498117).
- Hameleers, M. 2020. "Separating Truth from Lies: Comparing the Effects of News Media Literacy Interventions and Fact-Checkers in Response to Political Misinformation in the US and Netherlands." *Information, Communication & Society*, 1–17. doi:[10.1080/1369118X.2020.1764603](https://doi.org/10.1080/1369118X.2020.1764603).

- Hameleers, M. 2023. "Disinformation as a Context-Bound Phenomenon: Toward a Conceptual Clarification Integrating Actors, Intentions and Techniques of Creation and Dissemination." *Communication Theory* 33 (1): 1–10. doi: [10.1093/ct/qtac021](https://doi.org/10.1093/ct/qtac021).
- Hobbs, R., K. Donnelly, J. Friesem, and M. Moen. 2013. "Learning to Engage: How Positive Attitudes About the News, Media Literacy, and Video Production Contribute to Adolescent Civic Engagement." *Educational Media International* 50 (4): 231–246. doi:[10.1080/09523987.2013.862364](https://doi.org/10.1080/09523987.2013.862364).
- Hobbs, R., and S. McGee. 2014. "Teaching About Propaganda: An Examination of the Historical Roots of Media Literacy." *Journal of Media Literacy Education* 6 (2): 56–66. doi:[10.23860/jmle-6-2-5](https://doi.org/10.23860/jmle-6-2-5)
- Humprecht, E., L. Castro Herrero, S. Blassnig, M. Brüggemann, and S. Engesser. 2022. "Media Systems in the Digital Age: An Empirical Comparison of 30 Countries." *Journal of Communication* 72 (2): 145–164.
- Hwang, H., Z. Pan, and Y. Sun. 2008. "Influence of Hostile Media Perception on Willingness to Engage in Discursive Activities: An Examination of Mediating Role of Media Indignation." *Media Psychology* 11 (1): 76–97. doi:[10.1080/15213260701813454](https://doi.org/10.1080/15213260701813454).
- Jones-Jang, S. M., T. Mortensen, and J. Liu. 2021. "Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don." *American Behavioral Scientist* 65 (2): 371–388.
- Jungherr, A., and A. Rauchfleisch. 2022. "Negative Downstream Effects of Disinformation Discourse: Evidence from the US [Preprint]." *SocArXiv*. doi:[10.31235/osf.io/a3rzm](https://doi.org/10.31235/osf.io/a3rzm).
- Kohring, M., and J. Matthes. 2007. "Trust in News Media: Development and Validation of a Multidimensional Scale." *Communication Research* 34 (2): 231–252. doi:[10.1177/0093650206298071](https://doi.org/10.1177/0093650206298071).
- Levine, T. R. 2014. "Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection." *Journal of Language and Social Psychology* 33 (4): 378–392. doi:[10.1177/0261927X14535916](https://doi.org/10.1177/0261927X14535916).
- Levine, T. R. 2021. "Distrust, False Cues, and Below-Chance Deception Detection Accuracy: Commentary on Stel et al. (2020) and Further Reflections on (Un)Conscious Lie Detection from the Perspective of Truth-Default Theory." *Frontiers in Psychology* 12: 642359. doi:[10.3389/fpsyg.2021.642359](https://doi.org/10.3389/fpsyg.2021.642359).
- Levine, T. R., H. S. Park, and S. A. McCornack. 1999. "Accuracy in Detecting Truths and Lies: Documenting the "Veracity Effect"." *Communication Monographs* 66 (2): 125–144. doi:[10.1080/03637759909376468](https://doi.org/10.1080/03637759909376468).
- Levine, T. R., A. S. Shaw, and H. Shulman. 2010. "Assessing Deception Detection Accuracy with Dichotomous Truth–Lie Judgments and Continuous Scaling: Are People Really More Accurate When Honesty Is Scaled?." *Communication Research Reports* 27 (2): 112–122. doi:[10.1080/08824090903526638](https://doi.org/10.1080/08824090903526638).
- Luo, M., J. T. Hancock, and D. M. Markowitz. 2022. "Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues." *Communication Research* 009365022092132, doi:[10.1177/0093650220921321](https://doi.org/10.1177/0093650220921321).
- Lyons, B. A., V. Merola, and J. Reifler. 2020. "How Bad is the Fake News Problem?: The Role of Baseline Information in Public Perceptions." In *The Psychology of Fake News*, edited by R. Greifeneder, M. Jaffe, E. Newman, and N. Schwarz, 11–26. Abingdon: Routledge.
- Masullo, G. M., A. L. Curry, K. N. Whipple, and C. Murray. 2021. "The Story Behind the Story: Examining Transparency About the Journalistic Process and News Outlet Credibility." *Journalism Practice* 0 (0): 1–19. doi:[10.1080/17512786.2020.1870529](https://doi.org/10.1080/17512786.2020.1870529).
- Mcdougall, J. 2019. "Media Literacy Versus Fake News: Critical Thinking, Resilience and Civic Engagement." *Media Studies* 10 (19): 29–45. doi:[10.20901/ms.10.19.2](https://doi.org/10.20901/ms.10.19.2).
- McGrew, S., J. Breakstone, T. Ortega, M. Smith, and S. Wineburg. 2018. "Can Students Evaluate Online Sources? Learning from Assessments of Civic Online Reasoning." *Theory & Research in Social Education* 46 (2): 165–193. doi:[10.1080/00933104.2017.1416320](https://doi.org/10.1080/00933104.2017.1416320).
- Mihailidis, P. 2009. "Beyond Cynicism: Media Education and Civic Learning Outcomes in the University." *International Journal of Learning and Media* 1 (3): 19–31. doi:[10.1162/ijlm\\_a\\_00027](https://doi.org/10.1162/ijlm_a_00027)

- Mihailidis, P., and B. Foster. 2021. "The Cost of Disbelief: Fracturing News Ecosystems in an Age of Rampant Media Cynicism." *American Behavioral Scientist* 65 (4): 616–631. doi:[10.1177/0002764220978470](https://doi.org/10.1177/0002764220978470).
- Mihailidis, P., and B. Thevenin. 2013. "Media Literacy as a Core Competency for Engaged Citizenship in Participatory Democracy." *American Behavioral Scientist* 57 (11): 1611–1622. doi:[10.1177/0002764213489015](https://doi.org/10.1177/0002764213489015).
- Mihailidis, P., and S. Viotty. 2017. "Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in "Post-Fact"." *Society. American Behavioral Scientist* 61 (4): 441–454. doi:[10.1177/0002764217701217](https://doi.org/10.1177/0002764217701217).
- Newman, N., R. Fletcher, A. Schulz, S. Andi, C. T. Robertson, and R. K. Nielsen. 2022. *Reuters Institute Digital News Report 2022*. Reuters Institute for the study of Journalism.
- Nyhan, B., and J. Reifler. 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32 (2): 303–330. doi:[10.1007/s11109-010-9112-2](https://doi.org/10.1007/s11109-010-9112-2)
- Panizza, F., P. Ronzani, S. Mattavelli, T. Morisseau, C. Martini, and M. Motterlini. 2022. *Advised or Paid Way to get it Right. The Contribution of Fact-Checking Tips and Monetary Incentives to Spotting Scientific Disinformation*. doi:[10.21203/rs.3.rs-952649/v1](https://doi.org/10.21203/rs.3.rs-952649/v1).
- Pennycook, G., T. D. Cannon, and D. G. Rand. 2018. "Prior Exposure Increases Perceived Accuracy of Fake News." *Journal of Experimental Psychology: General* 147 (12): 1865–1880. doi:[10.1037/xge0000465](https://doi.org/10.1037/xge0000465).
- Pennycook, G., Z. Epstein, M. Mosleh, A. A. Arechar, D. Eckles, and D. G. Rand. 2021. "Shifting Attention to Accuracy Can Reduce Misinformation Online." *Nature* 592 (7855): 590–595. doi:[10.1038/s41586-021-03344-2](https://doi.org/10.1038/s41586-021-03344-2).
- Pennycook, G., J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand. 2020. "Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention." *Psychological Science* 31 (7): 770–780. doi:[10.1177/0956797620939054](https://doi.org/10.1177/0956797620939054).
- Pennycook, G., and D. G. Rand. 2019. "Lazy, not Biased: Susceptibility to Partisan Fake News is Better Explained by Lack of Reasoning Than by Motivated Reasoning." *Cognition* 188: 39–50. doi:[10.1016/j.cognition.2018.06.011](https://doi.org/10.1016/j.cognition.2018.06.011).
- Pennycook, G., and D. G. Rand. 2020. "Who Falls for Fake News? The Roles of Bullshit Receptivity, Overclaiming, Familiarity, and Analytic Thinking." *Journal of Personality* 88 (2): 185–200. doi:[10.1111/jopy.12476](https://doi.org/10.1111/jopy.12476).
- Petty, R. E., and W. B. G. Jarvis. 1996. *An individual differences perspective on assessing cognitive processes*.
- Roozenbeek, J., and S. van der Linden. 2019. "The Fake News Game: Actively Inoculating Against the Risk of Misinformation." *Journal of Risk Research* 22 (5): 570–580. doi:[10.1080/13669877.2018.1443491](https://doi.org/10.1080/13669877.2018.1443491).
- Soroka, S., P. Fournier, and L. Nir. 2019. "Cross-national Evidence of a Negativity Bias in Psychophysiological Reactions to News." *Proceedings of the National Academy of Sciences* 116 (38): 18888–18892. doi:[10.1073/pnas.1908369116](https://doi.org/10.1073/pnas.1908369116)
- Soroka, S., and S. McAdams. 2015. "News, Politics, and Negativity." *Political Communication* 32 (1): 1–22. doi:[10.1080/10584609.2014.881942](https://doi.org/10.1080/10584609.2014.881942).
- Southwell, B. G., E. A. Thorson, and L. Sheble. 2018. *Misinformation and Mass Audiences*. Austin, TX: University of Texas Press.
- Tsfati, Y., and J. N. Cappella. 2003. "Do People Watch What They Do Not Trust?: Exploring the Association Between News Media Skepticism and Exposure." *Communication Research* 30 (5): 504–529. doi:[10.1177/0093650203253371](https://doi.org/10.1177/0093650203253371).
- Tully, M., E. K. Vraga, and L. Bode. 2020. "Designing and Testing News Literacy Messages for Social Media." *Mass Communication and Society* 23 (1): 22–46. doi:[10.1080/15205436.2019.1604970](https://doi.org/10.1080/15205436.2019.1604970).
- Tully, M., E. K. Vraga, and A.-B. Smithson. 2020. "News Media Literacy, Perceptions of Bias, and Interpretation of News." *Journalism* 21 (2): 209–226. doi:[10.1177/1464884918805262](https://doi.org/10.1177/1464884918805262).
- Van Aelst, P., J. Strömbäck, T. Aalberg, F. Esser, C. de Vreese, J. Matthes, D. Hopmann, et al. 2017. "Political Communication in a High-Choice Media Environment: A Challenge for Democracy?" *Annals of the International Communication Association* 41 (1): 3–27. doi:[10.1080/23808985.2017.1288551](https://doi.org/10.1080/23808985.2017.1288551).

- van der Meer, T. G. L. A., and M. Hameleers. 2020. "Fighting Biased News Diets: Using News Media Literacy Interventions to Stimulate Online Cross-Cutting Media Exposure Patterns." *New Media & Society* 146144482094645. doi:[10.1177/1461444820946455](https://doi.org/10.1177/1461444820946455).
- van der Meer, T. G. L. A., and M. Hameleers. 2022. "I Knew It, the World is Falling Apart! Combatting a Confirmatory Negativity Bias in Audiences' News Selection Through News Media Literacy Interventions." *Digital Journalism* 0 (0): 1–20. doi:[10.1080/21670811.2021.2019074](https://doi.org/10.1080/21670811.2021.2019074).
- van der Meer, T. G. L. A., A. C. Kroon, P. Verhoeven, and J. Jonkman. 2019. "Mediatization and the Disproportionate Attention to Negative News: The Case of Airplane Crashes." *Journalism Studies* 20 (6): 783–803. doi:[10.1080/1461670X.2018.1423632](https://doi.org/10.1080/1461670X.2018.1423632).
- Van Duyn, E., and J. Collier. 2019. "Priming and Fake News: The Effects of Elite Discourse on Evaluations of News Media." *Mass Communication and Society* 22 (1): 29–48. doi:[10.1080/15205436.2018.1511807](https://doi.org/10.1080/15205436.2018.1511807).
- Vraga, E. K., and L. Bode. 2020. "Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation." *Political Communication* 37 (1): 136–144. doi:[10.1080/10584609.2020.1716500](https://doi.org/10.1080/10584609.2020.1716500).
- Vraga, E. K., L. Bode, and M. Tully. 2020. "Creating News Literacy Messages to Enhance Expert Corrections of Misinformation on Twitter." *Communication Research* 009365021989809. doi:[10.1177/0093650219898094](https://doi.org/10.1177/0093650219898094).
- Vraga, E. K., L. Bode, and M. Tully. 2022. "The Effects of a News Literacy Video and Real-Time Corrections to Video Misinformation Related to Sunscreen and Skin Cancer." *Health Communication* 37 (13): 1622–1630. doi:[10.1080/10410236.2021.1910165](https://doi.org/10.1080/10410236.2021.1910165).
- Vraga, E. K., and M. Tully. 2019. "Engaging with the Other Side: Using News Media Literacy Messages to Reduce Selective Exposure and Avoidance." *Journal of Information Technology & Politics* 16 (1): 77–86. doi:[10.1080/19331681.2019.1572565](https://doi.org/10.1080/19331681.2019.1572565).
- Vraga, E., M. Tully, and L. Bode. 2022. "Assessing the Relative Merits of News Literacy and Corrections in Responding to Misinformation on Twitter." *New Media & Society* 24 (10): 2354–2371. doi:[10.1177/1461444821998691](https://doi.org/10.1177/1461444821998691).
- Vraga, E. K., M. Tully, A. Maksl, S. Craft, and S. Ashley. 2021. "Theorizing News Literacy Behaviors." *Communication Theory* 31 (1): 1–21. doi:[10.1093/ct/qtaa005](https://doi.org/10.1093/ct/qtaa005).
- Vrij, A. 2008. *Detecting Lies and Deceit: Pitfalls and Opportunities*. Chichester: Wiley.
- Waisbord, S. 2018. "Truth is What Happens to News: On Journalism, Fake News, and Post-Truth." *Journalism Studies* 19 (13): 1866–1878. doi:[10.1080/1461670X.2018.1492881](https://doi.org/10.1080/1461670X.2018.1492881).