



UvA-DARE (Digital Academic Repository)

Beyond Belief Correction: Effects of the Truth Sandwich on Perceptions of Fact-checkers and Verification Intentions

Tulin, Marina; Hameleers, Michael ; de Vreese, Claes; Opgenhaffen, Michael; Wouters, Ferre

DOI

[10.1080/17512786.2024.2311311](https://doi.org/10.1080/17512786.2024.2311311)

Publication date

2025

Document Version

Final published version

Published in

Journalism Practice

License

CC BY-NC-ND

[Link to publication](#)

Citation for published version (APA):

Tulin, M., Hameleers, M., de Vreese, C., Opgenhaffen, M., & Wouters, F. (2025). Beyond Belief Correction: Effects of the Truth Sandwich on Perceptions of Fact-checkers and Verification Intentions. *Journalism Practice*, 19(11), 2576-2595.
<https://doi.org/10.1080/17512786.2024.2311311>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Beyond Belief Correction: Effects of the Truth Sandwich on Perceptions of Fact-checkers and Verification Intentions

Marina Tulin^a, Michael Hameleers^a, Claes de Vreese^a, Michaël Opgenhaffen^b and Ferre Wouters^b

^aAmsterdam School of Communication Research (ASCoR), University of Amsterdam, Amsterdam, Netherlands; ^bInstitute for Media Studies, KU Leuven, Leuven, Belgium

ABSTRACT

As mis- and disinformation may threaten democracy by fueling misperceptions, it is important to assess the effectiveness of journalistic interventions combating false information. This study aims to better understand how fact-checks relate to various outcomes relevant to audiences' resilience to false information. We randomly exposed 752 Dutch participants to fact-checks of a disputed health-related claim vs. no fact-check. The fact-checks either followed a classic format which repeated the false claim or followed a truth sandwich format wrapping the false claim in accurate information. While the truth sandwich was not effective in correcting false beliefs, it had indirect benefits. First, those who saw a truth sandwich perceived the *intentions* of fact-checkers more positively thinking that their intention was to inform rather than to manipulate or spread lies. Second, those who saw a truth sandwich showed the *least resistance* to reading subsequent fact-checks. For journalism practice this implies that different fact-check formats can be strategically employed to achieve desired outcomes. A more classic fact-check format might be preferable if the primary aim is to correct false beliefs, while the truth sandwich may be employed to reach more indirect and long-term aims like rebuilding confidence in fact-checkers or stimulating future verification behaviors.

ARTICLE HISTORY

Received 20 May 2023
Accepted 22 January 2024

KEYWORDS

Misinformation;
disinformation;
misperceptions; fact-check;
truth sandwich; belief
correction

Introduction

Mis- and disinformation are widely considered to be threatening our democracies. The spread of false information may contribute to the erosion of a common truth, uncertainty about reality, or the strengthening of socio-political cleavages (Bennett and Livingston 2018). Journalists are in a unique position to address this problem via a number of strategies, like fact-checking, providing corrective information and directing audiences towards more reliable sources (Sarelska and Jenkins 2023). These journalistic strategies seem to be successful: Recent meta-analyses show that fact-checks and corrective

CONTACT Marina Tulin  m.tulin@uva.nl

© 2024 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-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. 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.

information can help to correct false beliefs (e.g., Walter and Tukachinsky 2020). However, the effectiveness of fact-checks has mainly been conceptualized as the extent to which they are able to lower misperceptions, and it is therefore largely unclear what other benefits fact-checking might have. This is surprising because studies considering journalists' motivations to engage in fact-checking confirm that fact-checking practices are driven by broader professional motivations that go beyond belief correction, namely reaffirming journalists' role as political watchdogs and regaining professional status as objective truth-seekers for the service of the public (Graves, Nyhan, and Reifler 2016). In line with this, there is evidence that under certain circumstances, fact-checking can afford journalists more than only a better informed audience, namely increased media trust, reader self-efficacy and future news use (Pingree et al. 2018). Despite this initial evidence, we still know relatively little about the broader benefits of fact-checking, such as whether it enhances trust in fact-checkers, promotes critical thinking or stimulates future verification behaviors among readers. We also have very limited empirical evidence for a widely praised fact-check format, namely the so-called "truth sandwich" that wraps up a deceptive claim between two layers of factual information. It is not fully clear to what extent the truth sandwich is effective in correcting false beliefs and whether it is superior to other fact-check formats with regard to belief correction as well as other possible benefits.

The goal of this study is to better understand how fact-checking interventions relate to a number of outcomes relevant to audiences' resilience to false information. Together with journalists and fact-checking experts from the European Digital Media Observatory, a large European initiative against disinformation, we constructed fact-checks that varied in the location and prominence of false information. The fact-checks either wrapped the false claim in truthful information following a truth sandwich format, or they followed a classic format, which led with the false claim and debunked it. The fact-checks in classic format also varied with regard to the use of labels; they either explicitly mentioned labels that communicated a clear verdict on the falsehoods of verified statements or they avoided labels and left more room for the reader to draw their own conclusions. These variations sought to cover the most common formats of fact-check articles, which we aim to contrast against the truth sandwich format.

First we compared to what extent the truth sandwich vs. more classic fact-check formats are able to correct beliefs related to a false health claim. The aim is to address a theoretical conundrum: On the one hand, fact-checks need to address false claims in order to debunk them. On the other hand, repetition of false claims can enhance the credibility of false claims (Pillai and Fazio 2021). The alleged success of the truth sandwich is based on its ability to package incorrect information successfully, such that falsehoods get disarmed without backfiring (Lewandowsky and Cook 2020). However, to our knowledge there is currently no empirical basis for the claim that the truth sandwich is indeed more successful than classic fact-check formats.¹ Second, we looked beyond the direct effects on belief correction by exploring the extent to which these different fact-checks can stimulate critical thinking and trust in fact-checkers. Third, we considered whether the different fact-checks have an effect on the selection of subsequent fact-checks of other dubious claims.

Together, this experiment aims to explore the impact of the salience of references to falsehoods in fact-checking information and their effects on critical behaviors that help people to navigate other information in their newsfeed. As such, this study contributes to our

understanding of the benefits of reading fact-checks that go beyond belief correction by studying other key outcomes such as trust in fact-checkers, critical thinking, and future verification intentions. Our findings confirm that studying these outcomes matters, because different fact-checking formats diverge in their effects on these broader benefits of fact-checking.

Theory

Misinformation and Corrective Information

Misinformation can be understood as an umbrella-term to describe information that is not based on relevant expert knowledge and/or verified empirical evidence (Vraga and Bode 2020). In the literature on misinformation, a distinction between disinformation and misinformation is often made (e.g., Freelon and Wells 2020; Wardle and Derakhshan 2017). Misinformation refers to the unintentional or involuntary dissemination of false information, for example, due to a lack of expert knowledge or the inconsistency of evidence (Wardle and Derakhshan 2017). Disinformation, in contrast, pertains to the goal-directed dissemination of false information that is created with the intention to deceive recipients (Chadwick and Stanyer 2022; Hameleers 2023). In the political setting, the intention driving deceptive information may be to enhance cynicism, delegitimize established political actors, or increase societal cleavages to create momentum for alternative political movements and foreign influence (Freelon and Wells 2020). Although the distinction between mis- and disinformation is relevant when it comes to (legal) responsibilities, perceived consequences and perceptions as well as the intentions of actors to deceive are difficult to establish empirically — especially without contextual information on the drivers behind deceptive information campaigns (Hameleers 2023). Considering that intentions are difficult to determine, and that inaccurate information may have real consequences irrespective of intentions — we focus on misinformation in general in this article.

As misinformation has been associated with negative ramifications for democracy, different interventions to correct its impact have been introduced, among which fact-checks or corrective messages debunking falsehoods are the most prominent. Fact-checks are journalistic products — either from independent fact-checking platforms or news organizations — that check the factual basis of dubious claims (Uscinski and Butler 2013). Fact-checkers often rely on an argument-based style to refute or verify statements, and arrive at an overall verdict of the “degree of facticity” of statements (Graves and Amazeen 2019). Many scholars have indicated that fact-checks are effective in correcting misperceptions (e.g., Walter and Tukachinsky 2020). In addition, in a context in which falsehoods abound and alternative or counter-factual narratives compete for the audience’s attention, rigorous fact-checking by independent organizations remains an important journalistic practice (Amazeen 2015).

This is not to say that fact-checking efforts are without pitfalls or devoid of resistance from the audience. As explicated by Uscinski and Butler (2013), fact-checkers’ selection of claims to check are potentially informed by a bias. Specifically, fact-checkers do not rely on a representative sample of all (political) statements that were published in a given period, but rather decide to select and check statements that they deem untrustworthy

or dubious, for example, as they went viral online. Although such a bias may not be intentional, it may raise suspicion on the side of the audience. In light of this, extant research has shown that fact-checking is often avoided when it contradicts people's existing beliefs (e.g., Hameleers and van der Meer 2020). In a similar vein, in the bi-partisan setting of the US, partisans tend to selectively share corrective information that attacks opposed candidates or supports favored candidates (Shin and Thorson 2017). Supportive of this partisan or ideological bias in fact-checking perceptions, Republicans perceive that fact-checking is biased against their views, and that fact-checks disproportionately attack their party (Shin and Thorson 2017). Thus, although fact-checks have been regarded as effective in experimental studies, they may be subject to distrust and perceptions of bias by the audience. Specifically, although the mission statement of fact-checkers may be to offer a fact-based, independent, and neutral verdict of the truth value of (political) information (Amazeen 2015), the audience may not always perceive fact-checkers' intentions as honest and informed by balance and neutrality. In this context, this paper investigates the effectiveness of different formats of fact-checking while exploring to what extent these are effective beyond correcting misperceptions, which includes the association of dishonest intentions with the principles of fact-checking.

The Effectiveness of Fact-checks

Fact-checks typically check the veracity of dubious or prominent claims by testing them against empirical evidence and expert knowledge (Amazeen 2015). Most independent fact-checking organizations, such as *fullfact.org* in the UK, *PolitiFact.com* in the US, *Nieuwscheckers* in the Netherlands, or *Correctiv* in Germany, inform the public about the truthfulness of political statements by a verdict of the degree of facticity. Fact-checks use short and comprehensible messages with fact-based statements, which are used to arrive at a non-partisan evaluation of a message or statements (Lewandowsky et al. 2012). Using factual and short counterarguments after people have seen misinformation should result in recipients' acceptance of corrective information.

In line with this, numerous empirical studies — mostly based on survey experiments — have concluded that fact-checks are overall effective, seeing that exposure to fact-checking information reduces misperceptions (for meta-analyses, see: Walter and Murphy 2018; Walter et al. 2020). We could interpret the general effectiveness of fact-checks from the perspective of the Truth-Default-Theory (TDT) (Levine 2014). This theory holds that, across the board, people are more likely to accept the honesty of information than to question its trustworthiness and honesty. This truth bias has been confirmed empirically: Based on a meta-analysis, people are more likely to accurately judge accurate information as truthful than to rate false information as dishonest (Bond and DePaulo 2006). The TDT, however, also postulates that the activation of suspicion can break through the default state of the acceptance of truthfulness. More specifically, certain "trigger events" (Levine 2014), such as a lack of argument coherence or a strong deviation from a familiar reality (Luo, Hancock, and Markowitz 2022) can make people consider the possibility of being deceived (Clare and Levine 2019). Applied to misinformation and fact-checks, this means that people are generally more likely to trust false information than to question, at least as long as there is no

trigger event that might raise the suspicion that they are being deceived. However, the presentation of a fact-check could be considered a trigger event that raises suspicion about previously trusted misinformation. By flagging a discrepancy between previous information and reality, fact-checks emphasize the need to deviate from the truth-default. Therefore, exposure to a fact-check in response to misinformation should reduce misperceptions, and lower the truth value or credibility of statements based on misinformation. We therefore hypothesize:

H1: Participants exposed to a fact-check correcting misinformation are less likely to evaluate false claims as credible than participants who are not exposed to a fact-check.

The Role of Fact-Check Formats: The Truth Sandwich vs. Other Correction Styles

Fact-checks can be presented in different formats. Some fact-checks may use rating scales or labels to explicate the verdict, whereas other corrections offer a less clear verdict on the truthfulness of refuted claims (e.g., Amazeen et al. 2018). In addition, some fact-checks repeat false claims whereas others focus more on factually accurate information to avoid the prominence of falsehoods. These different decisions on how to fact-check may matter for its effectiveness, although we currently lack systematic empirical research that has explored the role of both explicit verdicts and the prominence of false information versus accurate content in fact-checks. We therefore ask: To what extent and how does varying the style of the correction (truth sandwich versus highlighting false claims) and the verdict (an explicit verdict versus no clear judgment) influence the effectiveness of a fact-check?

One key variation in different fact-check formats is the prevalence and placement of factual vs. inaccurate information. Here, we specifically aim to compare regular fact-checks that repeat the false claims in their header with the so-called “truth sandwich” that wraps up the deceptive claim between two layers of factual information (Kotz, Giese, and König 2022; König 2023). As indicated by extant research, the repetition of false claims can enhance their credibility (Pillai and Fazio 2021). More specifically, when people are exposed to the same falsehoods repeatedly, they may be perceived as familiar with existing cognitions and associations, which lowers the likelihood that suspicion is triggered.

To avoid this unintended byproduct of repetition, fact-checks may refrain from repeating false claims in their verdict and header. The truth sandwich can help to achieve this. By structuring the fact-check’s argument in a way that emphasizes factually accurate information, recipients may be directed towards truthful information instead of deceptive claims. Therefore, the truth sandwich could overall be more effective in correcting misperceptions. Despite this assumption, empirical research on the effectiveness of this strategy is scarce. Preliminary research — predominantly published as pre-prints at the time of writing, indicate that truth sandwiches are effective at reducing agreement with false statements (König 2023) but not more effective than regular refutations that repeat false claims (Kotz, Giese, and König 2022). Arguably, although truth sandwiches make lies less prevalent, its verdict might also be more ambiguous and more open to interpretation. As fact-checks may be effective because of their clear verdict and their unambiguous rating (Lewandowsky et al. 2012), the truth sandwich format may result in relatively

less clarity. Against this conflicting background, the following research question is introduced:

RQ1: Does the format of a truth-sandwich increase the *effectiveness* of fact-checks?

While the power of the truth sandwich allegedly lies in its ability to divert attention from misinformation, classic fact-check formats might be able to achieve the same via a different strategy: Avoid explicit labels of the fact-check's verdict and instead let readers draw their own conclusions based on the evidence presented. Because clear verdicts in the form of labels or rating scales draw the reader's attention to the false claim, fact-checks that avoid such labels might be more akin to the truth sandwich. We therefore study to what extent classic fact-checks that avoid using a clear verdict show similar effects to the truth sandwich.

Empirical evidence on the use of clear verdicts is currently inconclusive. A meta-analysis shows that explicit verdicts using a visual element, like rating scales, do work but are slightly less effective than fact-check articles that avoid them (Walter et al. 2020). There is also counter-evidence from several individual studies, like Clayton et al. (2020) who found that using "rated false" warning flags were more effective in refuting misinformation than general references to disputed claims. In line with this, Amazeen et al. (2018) found that the use of rating scales can be more effective than the absence of such clear labels. The inclusion of a rating scale resulted in stronger effects on corrected misperceptions. Moreover, fact-checks with a rating scale were more likely to be selected, and there were no negative backlash effects among partisans exposed to counter-attitudinal fact-checks with rating scales. Although these findings are optimistic, the effects of adding a rating scale or verdict are relatively small in these studies.

Evidence on the effects of ratings and explicit verdicts is thus overall inconclusive. In addition, most of this research is based on the partisan U.S. context, and focuses on political misinformation. We know relatively little about the extent to which these findings are transferable to other contexts, like the multi-party systems that are present in many European countries and whether the use of explicit labels has similar effects in the case of non-political information, like health claims, which we study in this paper. We therefore pose the following research question:

RQ2: Does the inclusion of an explicit label emphasizing the verdict of a fact-check increase the effectiveness of fact-checks?

Effects of Fact-checks Beyond Correcting Misperceptions

The Effects of Fact-checks on Trust in Truth-Seeking Institutions

Extant research on the effectiveness of fact-checks have mainly explored the effects of corrective information on the credibility of false claims or lowering misperceptions as a consequence of exposure to fact-checking information (e.g., Hameleers and van der Meer 2020; Thorson 2016). However, the effectiveness of fact-checks might (need to) be conceptualized more broadly. For fact-checks to work, fact-checkers and their truth-seeking practices need to be trusted by readers (Primig 2022), especially when the fact-checks address controversial political issues (Brandtzaeg and Følstad 2017). In line with this, journalists do not only engage in fact-checking to correct specific

beliefs, but also to reaffirm their professional values such as objectivity and their public service orientation (Graves, Nyhan, and Reifler 2016). Via fact-checking, journalists are able to demonstrate their good intentions to accurately inform the public and hold elites accountable. One study showed that fact-checks — when accompanied by a positive appraisal — are beneficial for higher media trust among readers, more self-efficacy and increased future news use (Pingree et al. 2018). Such benefits of fact-checks are critical in an era of factual relativism where the objective status of factual information and established media are under attack (Van Aelst et al. 2017). Established media outlets are frequently blamed for spreading “fake news” (Egelhofer and Lecheler 2019) and such delegitimizing labels can undermine citizens’ trust in factually accurate information (Van Duyn and Collier 2019). Fact-checkers who are often part of established journalistic outlets face similar attempts at delegitimization in an era where fake news accusations abound (Primig 2022). In line with this, Shin and Thorson (2017) point to a hostile media bias in partisans’ responses to fact-checkers. People are most likely to share fact-checks that discredit opposed candidates, and perceive fact-checking organizations as demonstrating a bias against their views. In this context, fact-checks are expected to be more effective when their aims reach beyond the correction of misperceptions. If successful, then fact-checking might ultimately serve to restore public trust in truth-seeking institutions.

Therefore, we specifically ask how the format of fact-checks may overcome distrust and resistance, especially when it comes to the perceived bias and intentions of fact-checkers as journalistic platforms. Arguably, reliance on the truth sandwich and the avoidance of explicit labels is more likely to overcome resistance from recipients whose attitudes align with the misinformation. Hence, as the verdict of such formats does not clearly attack the beliefs of recipients, the fact-check may be perceived as less hostile. Fact-checks that debunk statements by rating them as false or deceptive may be perceived as biased and misleading by recipients who supported the misinformed claims; the fact-check may be perceived as an attack on their beliefs. Despite these expectations, we lack empirical research on the effects of different fact-check formats on perceptions toward fact-checkers’ intentions, or overall levels of trust in factual information and expert knowledge.

We therefore introduce the following exploratory research questions:

What is the effect of different fact-checking formats (truth sandwich vs. classic formats) on trust in fact-checkers (**RQ3**) and perceived intentions of fact-checkers (**RQ4**)?

Restored or increased trust in fact-checkers might spill-over to other trust-seeking institutions like science or journalism, and the other way around. Trust in fact-checks, journalism and science are likely linked (Schäfer et al. 2018; Weingart and Guenther 2016; Pingree et al. 2018), and their connection matters for our study because we test fact-checks that verify a science-related claim. Trust in science depends on trust in the sources that communicate to the broader public about scientific issues, which is typically journalism and the news media (Weingart and Guenther 2016). Those who are most positive about science also rely heavily on the news media to learn more about scientific issues, while those who have more reservations about science also tend to be more critical of the news media (Schäfer et al. 2018). In addition, there is evidence for a positive effect of fact-checks on trust in the news media (Pingree et al. 2018).

In our study, we consider the possible spillover effects for trust in science because the fact-checks we test address a false health claim and draw on scientific explanations to debunk it. They also present scientific evidence to provide more accurate information. By drawing on scientific sources as the antidote to non-factual claims and providing rational and transparent scientific explanations, fact-checks might be able to strengthen trust in science. Because the truth sandwich judges false claims less harshly and instead highlights accurate (and in the case of our topic: non-controversial) scientific information, it might be particularly likely to produce positive spill-over effects for trust in science. The opposite could be true as well: A fact-check might fail to convince, perhaps because fact-checkers are seen as biased and selective in choosing their sources (Brandtzaeg and Følstad 2017), and this might have a negative spillover effect on trust in the scientific sources featured in the fact-check. We study this open question and the extent to which spill-over effects depend on the fact-check format by asking:

What is the effect of different fact-check formats on trust in science (RQ5)?

The Effects of Fact-check Formats on Media Literacy and Future Verification Behaviors

Finally, fact-checks may be more successful when they can instill critical media literacy skills among recipients, which can make them more resilient to misinformation in general. In an overloaded information ecology where false information spreads more widely than corrective information (Vosoughi, Roy, and Aral 2018), fact-checkers might not have the capacity to check all dubious claims. As such, it is desirable to promote more resilience to false information, and fact-checks might be able to teach the necessary skills for citizens to verify information and protect themselves against manipulation attempts. A recent study showed that fact-checks combined with media literacy messages are particularly effective in countering the negative effects of false information (Hamelers 2023).

While not all fact-checks include explicit media literacy messages, an important principle of fact-checking, which is also key to media literacy among citizens, is that it distinguishes facts from opinions or lies. As such, fact-checkers are exceptionally transparent in showing the process of evaluating claims (e.g., see the International Fact-checking Network Code of Principles, Poynter 2023) as well as tracking down original sources, and verifying evidence (Mena 2019). A fact-check article typically does not only provide an assessment, but also a backstory explaining exactly how they arrived at their conclusions. This can include explanations of why a claim is false or misleading and how the false evidence was fabricated or taken out of context. Whether explicitly or implicitly, fact-checks often communicate how falsehoods can be spotted, and in this way, they might be able to protect readers against future manipulation attempts. In support of this, a recent study found that citizens who learned about common fact-checking practices, such as comparing different sources, were more accurate at discerning reliable from unreliable sources afterwards (Panizza et al. 2021). Building on this preliminary evidence, our study seeks to find out whether different fact-check formats are able to instill media literacy among citizens. We are also interested in testing whether and which fact-

check formats have the most potential to motivate citizens to engage in verification behaviors. We therefore ask:

What is the effect of different fact-check formats on media literacy (**RQ6**) and future verification behaviors? (**RQ7**)

Methods

This study has received ethical approval from the Ethical Review Board of the University of Amsterdam under project number 2022-PCJ-15378. The ethics application was evaluated based on a detailed project description, informed consent form, and data management plan. The research questions, hypotheses, design, sampling, and sample size were pre-registered with the Open Science Framework (link: https://osf.io/cs4tq/?view_only=4fa6a2dd199c45f887a43c12a2ebd1b2). With the exception of two small deviations, we closely followed the pre-registration: First, the order of the research questions was adjusted to accommodate a better flow in the theory section. Second, in the pre-registration we mention that we measure evaluations of three claims: one false claim, one partly false claim, and one true claim. For space constraints, however, this paper focuses on analyses of the false claim evaluations only because these align best with the theoretical contributions. Results for the other two claims can be found in the Appendix (see Tables A1 and A2). Finally, the pre-registration also mentions RQ8 on resistance to fact-checking, which is being addressed in a separate paper.

Sample and Procedure

The data for this study were collected in an online survey experiment among Dutch citizens. A total of 1310 individuals were invited to take the survey and 752 individuals completed the survey, which means that the response rate was 57.4%. The sample approximated the Dutch population in terms of gender (52% women), education (38% had a college or university degree) and age (36% were <40 years old, 39% were 40–64 years old and 25% were 65 years or older).

Participants first filled out some background information and were then randomly exposed to a fact-check of a disputed health-related claim related to weight loss. To ensure that our fact-checks accurately reflected the type of information that readers might come across in the real world, we selected an actual misinformation statement related to dieting. The false statement claimed that the order in which individuals eat their meals has a positive effect on weight loss, irrespective of the amount of calories they consume. We selected a health claim that was of broad interest but not directly linked to strong ideological biases. In the control condition participants did not see a

Table 1. Overview of all conditions in the experiment.

		Fact-check style			
		Label: Yes	Label: No	Truth sandwich	No fact-check (control)
Factuality level	False	Condition 1	Condition 2	Condition 5	Condition 0
	Mostly false	Condition 3	Condition 4		

Note: Conditions 1–4 follow a classic fact-check format that repeats the false claim in the header and introduction.

fact-check. In the experimental conditions, participants read a fact-check debunking the false claim. The fact-check followed either a classic format varying the label and level of factuality (Conditions 1–4) or a truth sandwich format (Condition 5). In an effort to make the fact-checks more realistic and to enhance the robustness of our findings, the classic fact-check formats included variations on the use of labels that communicate a clear verdict (yes vs. no) as well as level of factuality such that the claims were either judged as “false” or “completely false” (see Table 1). The variation on level of factuality in particular was done for external validity purposes, trying to mimic the actual existence and complicated nature of untruthfulness in misinformation.

To further increase external validity, we constructed the fact-checks in collaboration with professional fact-checkers from a European fact-checking organization, namely Nieuwscheckers. The fact-checks were formulated and formatted to resemble the style of this fact-checking organization. As a control question, we asked participants at the end of the survey whether they had been familiar with the fact-check organization before participating in this study (yes, no, unsure). A minority of respondents (14%) said that they were familiar with Nieuwscheckers and this proportion was evenly

Table 2. Descriptive statistics.

	%	Mean	SD	Min	Max
Gender					
Man	48				
Woman	52				
Prefer not to say	0				
Education					
Basic education	2				
Pre-vocational secondary education	15				
Pre-university education	12				
Vocational education	32				
College	27				
University	11				
Other	1				
Age (years)					
16–19	2				
20–39	34				
40–64	39				
65–79	25				
80+	1				
Familiarity					
Yes	14				
No	78				
Unsure	7				
Rating of false claim		2.75	1.08	1	5
Trust in fact-checkers		2.74	.84	1	5
Perceived intention of fact-checker					
Positive intentions (composite)		4.77	1.02	1	7
Educate		4.44	1.48	1	7
Inform		5.18	1.38	1	7
Tell the truth		4.74	1.47	1	7
Manipulate		3.41	1.69	1	7
Spread lies		3.12	1.66	1	7
Stimulate critical thinking		4.77	1.42	1	7
Trust in science		4.45	1.15	1	7
Self-perceived media literacy		4.61	1.05	1	7
Selection of subsequent fact-check					
Related claim	47				
Unrelated claim	41				

spread across conditions ($\chi^2(10, n = 752) = 8.92, p = 0.540$). Accordingly, controlling for this variable in our main analyses did not change the results.

Measures

This study explores a number of dependent variables predicted by the type of fact-check that participants saw in the experimental conditions as compared to the control condition. Below we describe the dependent variables per hypothesis and research question. Table 2 presents descriptive statistics of sample characteristics and all dependent variables included in this study. Unless otherwise specified, all dependent variables are based on composite scores of relevant battery questions, which were created by averaging across items scores belonging to the same battery.

Rating of false claims (H1, RQ1, RQ2) is measured as the mean response to a five-item battery capturing the perceived credibility, accuracy, reliability, bias and completeness of the false health claim measured on a five-point scale (Cronbach's $\alpha = .89$).

Trust in fact-checkers (RQ3) is measured as the response to a five-item battery capturing how honest, biased, complete, accurate and trustworthy the fact-checker is perceived to be (measured on a five-point scale; Cronbach's $\alpha = .82$).

Perceived intentions of fact-checkers (RQ4) is measured as the degree of agreement with six statements (1 = fully disagree; 7 = fully agree). The statements read "The intention of the fact-checkers was to (1) educate, (2) inform, (3) tell the truth, (4) manipulate, (5) spread lies, (6) stimulate critical thinking". We present results for these six measures separately as well as combined under a composite measure. The composite measure was obtained by first reverse-coding responses to the items "manipulate" and "spread lies", and then averaging across responses to all six measures (Cronbach's $\alpha = .75$).

Trust in science (RQ5) is measured as the degree of agreement with ten statements (1 = fully disagree; 7 = fully agree; Cronbach's $\alpha = .90$), for example "I find it important to be informed about science and research".

Self-perceived media literacy (RQ6) is measured as the degree of agreement with four statements (1 = fully disagree; 7 = fully agree; Cronbach's $\alpha = .80$), for example "I find it easy to distinguish between factually correct and incorrect information".

Future verification behaviors (RQ7) captured whether and which fact-check participants chose to read after being presented with two news items at the end of the study. One news item was on a related topic, namely red wine and abdominal fat, while the other covered an unrelated topic, namely asylum seekers. Respondents selected whether they wanted to read fact-checks for (a) one, (b) both or (c) neither of the topics.

Results

To test Hypothesis 1 on whether exposure to a fact-check makes individuals evaluate false claims less positively, we conducted OLS regression analyses with fact-check conditions as independent variable and claim credibility as the dependent variable.² Specifically, we tested evaluations of the false claim that "The order of eating a meal has an effect on your weight". Results in Table 3 show the unstandardized and standardized regression coefficients as well as their confidence intervals per condition. ANOVA results show that the overall model is statistically significant meaning that the different fact-check

Table 3. Results of regression analyses where claim credibility is predicted by fact-check formats.

	Rating of false claim	
	B [95% CI]	Beta [95% CI]
Control condition: no fact-check	Ref.	Ref.
Classic fact-check formats		
Condition 1: label [yes] & factuality [false]	-0.26+ [-0.53, 0.00]	-0.09+ [-0.18, 0.00]
Condition 2: label [no] & factuality [false]	-0.05 [-0.31, 0.22]	-0.02 [-0.11, 0.07]
Condition 3: label [yes] & factuality [mostly false]	-0.11 [-0.37, 0.16]	-0.04 [-0.13, 0.05]
Condition 4: label [no] & factuality [mostly false]	-0.39** [-0.66, -0.13]	-0.13** [-0.23, -0.04]
Truth sandwich (Condition 5)	0.16 [-0.10, 0.42]	0.06 [-0.03, 0.15]
Constant	2.86*** [2.67, 3.04]	0.00 [-0.07, 0.07]
R^2	.028	
Observations	752	

Note: 95% confidence intervals in square brackets. + $p < 0.10$, ** $p < 0.01$, *** $p < 0.001$.

formats have an effect on ratings of the false claim ($F(5,746) = 4.29, p < .001$). Overall, the more classic fact-check formats were more successful at reducing false beliefs than the truth sandwich, seeing that the regression coefficients of the classic formats are negative across the board when compared to the control condition. The truth sandwich did not lead to a reduction in false beliefs. We observe that a classic fact-check that used no label and judged the claim as mostly false (Condition 4) was the most successful at reducing the credibility of the false claim ($B = -0.39, p = .004$), with a small effect size ($beta = -.13$). Another classic format that emerged as successful was the fact-check that did use an explicit label and judged the claim as completely false (Condition 1). This fact-check reduced the credibility of the false claim with a small effect size ($beta = -.09$), even though it failed to reach conventional levels of statistical significance by a very small margin ($B = -0.26, p = .051$).

Hypothesis 1 is thus supported for two out of five fact-check formats. Importantly, the hypothesis was not confirmed for the truth sandwich, seeing that the truth sandwich failed to lower agreement with the false claim. This also provides a partial answer to RQ1 which asked whether the truth sandwich format made fact-checks more effective. The results so far show that this is not the case for credibility ratings. However, when it comes to other dependent variables, the truth sandwich did stand out positively as will be shown in subsequent analyses (see for example results for RQ4).

To test RQ2 about the effects of including an explicit verdict in the form of a label, we computed a new variable that captured whether a fact-check label was present or not. To this end, we pooled Conditions 1 and 3 because they did include a label. Conditions 2, 4 and 5 were pooled because they did not include a label. Table 4 shows the results of including vs. not including a label in a fact-check as compared to the control condition. The inclusion of labels did not have a statistically significant effect on the credibility of the false claim (see Table 4).

In relation to RQ2, these results suggest that including a label does not make fact-checks more effective across the board. Rather, as seen in Table 3, we saw the largest

Table 4. Claim credibility predicted by whether the fact-check includes a label.

	Rating of false claim	
	B [95% CI]	Beta [95% CI]
Control condition: no fact-check	Ref.	
Fact-check label: yes	-0.18 [-0.42, 0.05]	-0.08 [-0.18, 0.02]
Fact-check label: no	-0.09 [-0.30, 0.13]	-0.04 [-0.14, 0.06]
Constant	2.86*** [2.67, 3.04]	0.00 [-0.07, 0.07]
R ²	.004	
Observations	752	

Note: 95% confidence intervals in square brackets. ****p* < 0.001.

effect in a fact-check that told a more complex story about the conditions under which the claim was true or false and that avoided using a clear label to communicate its verdict (Condition 4). While in this case a clear label was not helpful, it seems that the level of factuality that is communicated in the fact-check matters for whether a fact-check benefits from a clear label or not. Credibility ratings did reduce when the fact-check used a clear label to communicate that the claim was completely false (Condition 1, Table 3). This seems to suggest that whether labels are effective or not depends on the level of factuality of the claim.

RQ3 asked whether the fact-check formats have a significant effect on trust in fact-checkers. This did not turn out to be the case. However, the different fact-check

Table 5. Results of regression analyses where perceived intentions of fact-checkers are predicted by fact-check formats.

	(1) Positive intentions (composite)	(2) Educate	(3) Inform	(4) Tell the truth	(5) Manipulate	(6) Spread lies	(7) Stimulate critical thinking
Condition 5: truth sandwich	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Condition 1: label [yes] & factuality [false]	-0.26* [-0.51, -0.00]	-0.27 [-0.64, 0.09]	-0.41* [-0.75, -0.06]	-0.04 [-0.41, 0.32]	0.31 [-0.11, 0.72]	0.40+ [-0.01, 0.82]	-0.10 [-0.45, 0.25]
Condition 2: label [no] & factuality [false]	(-.10)* -0.29* [-0.53, -0.04]	(-.07) -0.39* [-0.75, -0.03]	(-.11)* -0.37* [-0.71, -0.04]	(-.01) -0.21 [-0.57, 0.16]	(.07) 0.26 [-0.15, 0.68]	(.10)+ 0.34 [-0.07, 0.75]	(-.03) -0.14 [-0.49, 0.20]
Condition 3: label [yes] & factuality [mostly false]	(-.11)* -0.08 [-0.33, 0.17]	(-.11)* -0.21 [-0.58, 0.15]	(-.11)* -0.15 [-0.49, 0.19]	(-.06) -0.01 [-0.37, 0.36]	(.06) 0.11 [-0.30, 0.53]	(.08) 0.25 [-0.16, 0.66]	(-.04) 0.26 [-0.08, 0.61]
Condition 4: label [no] & factuality [mostly false]	(-.03) -0.34** [-0.59, -0.09]	(-.06) -0.27 [-0.64, 0.09]	(-.04) -0.39* [-0.73, -0.05]	(-.00) -0.38* [-0.74, -0.02]	(.03) 0.30 [-0.11, 0.72]	(.06) 0.37+ [-0.05, 0.78]	(.07) -0.30+ [-0.65, 0.05]
Constant	(-.13)** 4.95*** [4.78, 5.13]	(-.07) 4.67*** [4.41, 4.92]	(-.11)* 5.44*** [5.20, 5.68]	(-.10)* 4.86*** [4.61, 5.11]	(.07) 3.22*** [2.92, 3.51]	(.09)+ 2.85*** [2.57, 3.14]	(-.08)+ 4.82*** [4.58, 5.07]
R ²	.02*	.01	.01	.01	.01	.01	.02*
Observations	624	624	624	624	624	624	624

Note: Estimates are unstandardized regression coefficients. 95% confidence intervals in square brackets and standardized regression coefficients in parentheses. +*p* < 0.10, **p* < 0.05, ***p* < 0.01, ****p* < 0.001.

formats did affect the perceived intentions of fact-checkers (RQ4, see Table 5). The truth sandwich stands out as the format that was perceived to be written with the most positive intentions. When considering all positive intentions combined (Table 5, Model 1), we observe that the truth sandwich is perceived significantly more positively than three of the four classic fact-check formats, namely Conditions 1, 2, and 4. The effect sizes for these comparisons were small (*betas* ranged from .10 to .13).

When considering the six intentions separately, the truth sandwich was seen as intending to educate more than the classic fact-check format in Condition 2 and to inform more than the classic formats in Conditions 1, 2 and 4. It was also seen as intending to tell the truth more so than Condition 4. Interestingly, the truth sandwich was also seen as less likely to spread lies as compared to Condition 1 and 4, even though these effects are only significant at the $p < .10$ level.

With regard to RQ5 about media literacy, we do not find that fact-check formats affected self-perceived media literacy skills. Similarly, we do not find that trust in science was affected by reading any of the fact-checks (RQ6).

RQ7 considered whether the fact-check format predicted future verification behaviors. Respondents could opt to read a fact-check on the same topic as the initial fact-check, or a different, unrelated topic or both. 35% of the sample opted to not read a subsequent fact-check, while 42% selected one, and 23% selected both fact-checks. Model 1 in Table 6 shows the likelihood of selecting at least one of the two fact-checks. The coefficients for all experimental conditions are negative, suggesting that reading any fact-check makes individuals less likely to select a subsequent fact-check. Especially when individuals read fact-checks that did not include a clear label

Table 6. Results of regression analyses where selection of subsequent fact-checks is predicted by initial fact-check formats.

	(1) At least one fact-check	(2) Same topic fact-check	(3) Other topic fact-check
Control condition: no fact-check	Ref.	Ref.	Ref.
Condition 1: label [yes] & factuality [false]	-0.05 [-0.17, 0.07] (-0.04)	-0.05 [-0.18, 0.07] (-0.04)	-0.13* [-0.25, -0.00] (-0.09*)
Condition 2: label [no] & factuality [false]	-0.12* [-0.24, -0.00] (-0.09*)	-0.07 [-0.19, 0.06] (-0.05)	-0.18** [-0.30, -0.06] (-0.14**)
Condition 3: label [yes] & factuality [mostly false]	-0.01 [-0.12, 0.11] (-0.01)	-0.01 [-0.14, 0.11] (-0.01)	-0.09 [-0.21, 0.03] (-0.07)
Condition 4: label [no] & factuality [mostly false]	-0.13* [-0.25, -0.02] (-0.10*)	-0.09 [-0.21, 0.04] (-0.06)	-0.20** [-0.32, -0.08] (-0.15**)
Condition 5: truth sandwich	-0.04 [-0.16, 0.07] (-0.03)	-0.02 [-0.14, 0.11] (-0.01)	-0.12* [-0.24, -0.00] (-0.09*)
Constant	0.71*** [0.63, 0.79]	0.51*** [0.42, 0.59]	0.53*** [0.45, 0.62]
R^2	.02	.004	.02*
Observations	752	752	752

Note: Estimates are unstandardized regression coefficients. 95% confidence intervals in square brackets and standardized regression coefficients in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

(Condition 2 and 4), participants were particularly reluctant to read a subsequent fact-check and this effect is statistically significant. When the fact-check did include a label (Conditions 1 and 3) or when the fact-check was in the truth sandwich, then the coefficients also point in the negative direction, however these effects were not significant. In other words, including labels or using the truth sandwich format did not fuel resistance to reading subsequent fact-checks.

To further investigate this phenomenon, we split the analyses by fact-check topic. Model 2 in [Table 6](#) shows that none of the fact-check formats created resistance to reading a fact-check on the same topic. However, there was general resistance to reading a fact-check on a different topic (see Model 3 in [Table 6](#)). When it comes to a different topic, the fact-checks without a clear label raised the most resistance.

Discussion

Considering that exposure to mis- and disinformation can enhance polarization, increase cynicism about established institutions, or cultivate misperceptions about conventional knowledge (e.g., Bennett and Livingston 2018), it is important to investigate how false claims can be debunked. Yet, we know markedly little about which formats of corrective information are most effective, and whether fact-checks can enhance resilience toward mis- and disinformation beyond correcting factual beliefs. At the same time, different scholars have argued that fact-checkers should do more than correcting misperceptions, for example, by enhancing critical verification skills and trust in journalism (e.g., Pingree et al. 2018). To offer new insights into the effectiveness of fact-checks beyond belief correction, this paper explored to what extent different formats of fact-checking (a classical refutation with and without specific labels versus a truth sandwich) impacted misperceptions as well as trust, perceived intentions, media literacy and future verification efforts.

Our main findings indicate that classical fact-checking formats were effective in correcting factual misperceptions, which confirms the consensus in experimental research that has dealt with similar questions (e.g., Hameleers and van der Meer 2020; Nyhan et al. 2020). However, we found important differences when we look at different indicators of effectiveness. Although we currently lack empirical evidence on the effectiveness of the truth sandwich, our findings suggest that it is less effective than traditional fact-checks in correcting factual misperceptions. Yet, the truth sandwich outperformed the traditional fact-check when it comes to perceived fact-checking intentions and resistance to additional fact-checks. We can explain these differential findings based on the different roles of facticity and prominence of false claims across fact-checking formats. The format of the truth sandwich makes the debunked claims less explicit, and focuses on the correct depiction of reality instead of discrediting the false statements (Kotz, Giese, and König 2022; König 2023). The traditional fact-check, in contrast, repeats the false statements and offers various arguments why this statement was false and not based on evidence, herewith offering a more concrete recommendation on the falsehood of the debunked statement.

In other words, the traditional fact-check format more explicitly and concretely triggers suspicion and herewith directly informs recipients on the falsehood of claims — potentially breaking through the truth bias or truth-default of information processing (Levine 2014).

Although such a direct trigger of suspicion may be effective in correcting factual misperceptions, it can also be regarded as an attack on people's prior beliefs (Thorson 2016) or a restriction on people's freedom to arrive at their own judgements. This may explain why the traditional format had a less positive effect on the perceived intentions of the corrective information and the motivation to select additional corrective information. The truth sandwich, in contrast, did not push a strong verdict on the untruthfulness of the health-related claims, and may thus be perceived as a suggestion and aid in the processing of information, whilst maintaining people's freedom to make their own judgements.

As a key theoretical implication, our paper offers first evidence of the effectiveness of using different formats of fact-checking when it comes to the positioning and emphasis of the debunked false statements, suggesting that the presumed positive influence of not repeating false claims but wrapping them up in layers of truth is less effective in correcting factual misperceptions. We also contribute to the literature by suggesting that effectiveness of fact-checks should not be equated with the ability of corrective information to lower misperceptions directly after exposure. Rather, the indirect aim of fact-checks may be to educate audiences to become more resilient toward mis- and disinformation, or to enhance trust in factually accurate information (Graves, Nyhan, and Reifler 2016).

Our findings have important implications for journalism practice. We show that different formats of fact-checking have different advantages and disadvantages, which means that the format of fact-checking may be selected and targeted depending on the intended aims of fact-checkers and journalists. If the aim is to quickly and repeatedly debunk false statements, for example in the midst of the rapid dissemination of false claims amidst a health-related crisis, a classical fact-check approach in which the false claim is emphasized and clearly debunked with labels may be the best strategy. However, if the goal is to enhance critical verification skills among the audience, and stimulate the use of corrective information on a longer-term base, the truth sandwich may be a suitable format of fact-checking. Arguably, the strengths of the different formats can be exploited by varying the presentation of corrective information, for example, by combining short traditional fact-checks directly responding to individual claims with longer background truth sandwich fact-checks that deal with a theme or pressing issue in its entirety.

Despite these implications, our study has a number of limitations. First of all, similar to existing fact-check experiments, we use an immediate measure of outcome variables without taking into account the longer-term effects of corrective information. While viewing fact-checkers' intentions more positively might indicate positive long-term effects of the truth sandwich, a one-wave experiment does not allow us to empirically demonstrate any cumulative positive effects that might occur over time. Second, we focus on mis- and disinformation on a health-related issue. The outcomes may be different for more polarizing political issues, where the role of prior beliefs and confirmation biases is arguably more decisive for the effectiveness of fact-checking (see e.g., Thorson 2016). Finally, this study was conducted in a single country and it is unclear to what extent the findings would generalize to other populations. Despite these limitations, we have taken great care to design a study that maximizes ecological validity by closely collaborating with fact-checking professionals from two prominent fact-checking organizations in the country studied. While more research is needed to understand the generalizability of our findings, we believe that our study provides an important starting point for

understanding the unique benefits and drawbacks of the truth sandwich format in fighting the negative consequences of mis- and disinformation.

Conclusion

Despite receiving wide praise in journalism practice, the truth sandwich does not have unequivocal empirical support as a positive impact on belief correction. At the same time, there does not seem to be a large disadvantage to using the truth sandwich format and it does have several advantages that go beyond belief correction. Readers of the truth sandwich assess the fact-checker's intentions more positively, and consider them to write the fact-check with the aim to inform, to educate or to tell the truth. Because the truth sandwich lacks a concrete judgment, it might be considered more similar to other journalistic pieces and be perceived more positively as a result. Unlike more classic fact-check formats that repeat the false claim or highlight the verdict, the truth sandwich does not trigger resistance to reading subsequent fact-checks. Depending on the aim, the truth sandwich can thus be considered a viable alternative to a more classic fact-check format.

Notes

1. At the time of writing, the only empirical studies on the truth sandwich are published as pre-prints and they do not support the claim that the truth sandwich is superior to more classic fact-check formats (Kotz, Giese, and König 2022; König 2023).
2. To test the robustness of our findings, we also performed ANOVAs and post-hoc tests for each of the regression analyses and the pattern of results is identical.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This paper has been written as part of the BENEDMO project which has received funding from the European Union under Grant Agreement number INEA/CEF/ICT/A2020/2381738.

References

- Amazeen, Michelle A. 2015. "Revisiting the Epistemology of Fact-Checking." *Critical Review* 27 (1): 1–22. doi:10.1080/08913811.2014.993890
- Amazeen, Michelle A., Emily Thorson, Ashley Muddiman, and Lucas Graves. 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. doi:10.1177/1077699016678186
- Bennett, W. Lance, and Steven 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
- Bond, Charles F., and Bella M. DePaulo. 2006. "Accuracy of Deception Judgments." *Personality and Social Psychology Review* 10 (3): 214–234. doi:10.1207/s15327957pspr1003_2

- Brandtzaeg, Petter Bae, and Asbjørn Følstad. 2017. "Trust and Distrust in Online Fact-Checking Services." *Communications of the ACM* 60 (9): 65–71. doi:10.1145/3122803
- Chadwick, Andrew, and James Stanier. 2022. "Deception as a Bridging Concept in the Study of Disinformation, Misinformation, and Misperceptions: Toward a Holistic Framework." *Communication Theory* 32 (1): 1–24. doi:10.1093/ct/qtab019
- Clare, David D., and Timothy 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
- Clayton, Katherine, Spencer Blair, Jonathan A. Busam, Samuel Forstner, John Glance, Guy Green, Anna Kawata, et al. 2020. "Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media." *Political Behavior* 42 (4): 1073–1095. doi:10.1007/s11109-019-09533-0
- Egelhofer, Jana Laura, and Sophie 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
- Freelon, Deen, and Chris Wells. 2020. "Disinformation as Political Communication." *Political Communication* 37 (2): 145–156. doi:10.1080/10584609.2020.1723755
- Graves, Lucas, and Michelle A. Amazeen. 2019. "Fact-Checking as Idea and Practice in Journalism." Oxford Research Encyclopedia of Communication. <https://oxfordre.com/communication/abstract/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-808>.
- Graves, Lucas, Brendan Nyhan, and Jason Reifler. 2016. "Understanding Innovations in Journalistic Practice: A Field Experiment Examining Motivations for Fact-Checking." *Journal of Communication* 66 (1): 102–138. doi:10.1111/jcom.12198
- Hameleers, Michael. 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
- Hameleers, Michael, and Toni G. L. A. van der Meer. 2020. "Misinformation and Polarization in a High-Choice Media Environment: How Effective Are Political Fact-Checkers?" *Communication Research* 47 (2): 227–250. doi:10.1177/0093650218819671
- König, Laura M. 2023. "Debunking Nutrition Myths: An Experimental Test of the 'Truth Sandwich' Text Format." *British Journal of Health Psychology* 28 (4): 1000–1010. doi:10.1111/bjhp.12665
- Kotz, Johannes, Helge Giese, and Laura M. König. 2022. "How to Debunk Misinformation? An Experimental Online Study Investigating Text Structures and Headline Formats." August. <https://psycharchives.org/en/item/07ff10f2-0b9a-4090-b8dc-4b3af0c6ba51>.
- Levine, Timothy 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
- Lewandowsky, Stephan, and John Cook. 2020. "The Debunking Handbook 2020." <https://sks.to/db2020>.
- Lewandowsky, Stephan, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. 2012. "Misinformation and Its Correction." *Psychological Science in the Public Interest* 13 (3): 106–131. doi:10.1177/1529100612451018
- Luo, Mufan, Jeffrey T. Hancock, and David M. Markowitz. 2022. "Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues." *Communication Research* 49 (2): 171–195. doi:10.1177/0093650220921321
- Mena, Paul. 2019. "Principles and Boundaries of Fact-Checking: Journalists' Perceptions." *Journalism Practice* 13 (6): 657–672. doi:10.1080/17512786.2018.1547655
- Nyhan, Brendan, Ethan Porter, Jason Reifler, and Thomas J. Wood. 2020. "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. doi:10.1007/s11109-019-09528-x
- Panizza, Folco, Piero Ronzani, Simone Mattavelli, Tiffany Morisseau, Carlo Martini, and Matteo Motterlini. 2021. "Advised or Paid Way to Get It Right. The Contribution of Fact-Checking Tips and Monetary Incentives to Spotting Scientific Disinformation." Preprint. In Review. doi:10.21203/rs.3.rs-952649/v1

- Pillai, Raunak M., and Lisa K. Fazio. 2021. "The Effects of Repeating False and Misleading Information on Belief." *WIREs Cognitive Science* 12 (6): e1573. doi:10.1002/wcs.1573
- Pingree, Raymond J., Brian Watson, Mingxiao Sui, Kathleen Searles, Nathan P. Kalmoe, Joshua P. Darr, Martina Santia, and Kirill Bryanov. 2018. "Checking Facts and Fighting Back: Why Journalists Should Defend Their Profession." *PLoS One* 13 (12): e0208600. doi:10.1371/journal.pone.0208600
- Poynter. 2023. "International Fact-Checking Network Code of Principles." <https://www.ifcncodeofprinciples.poynter.org/>.
- Primig, Florian. 2022. "The Influence of Media Trust and Normative Role Expectations on the Credibility of Fact Checkers." *Journalism Practice* 1–21. doi:10.1080/17512786.2022.2080102
- Sarelska, Darina, and Joy Jenkins. 2023. "Truth on Demand: Influences on How Journalists in Italy, Spain, and Bulgaria Responded to Covid-19 Misinformation and Disinformation." *Journalism Practice* 17 (10): 2178–2196. doi:10.1080/17512786.2022.2153075
- Schäfer, Mike S., Julia Tobias Füchslin, Silje Metag, and Adrian Rauchfleisch. Kristiansen. 2018. "The Different Audiences of Science Communication: A Segmentation Analysis of the Swiss Population's Perceptions of Science and Their Information and Media Use Patterns." *Public Understanding of Science* 27 (7): 836–856. doi:10.1177/0963662517752886
- Shin, Jieun, and Kjerstin Thorson. 2017. "Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media." *Journal of Communication* 67 (2): 233–255. doi:10.1111/jcom.12284
- Thorson, Emily. 2016. "Belief Echoes: The Persistent Effects of Corrected Misinformation." *Political Communication* 33 (3): 460–480. doi:10.1080/10584609.2015.1102187
- Uscinski, Joseph E., and Ryden W. Butler. 2013. "The Epistemology of Fact Checking." *Critical Review* 25 (2): 162–180. doi:10.1080/08913811.2013.843872
- Van Aelst, Peter, Jesper Strömbäck, Toril Aalberg, Frank Esser, Claes de Vreese, Jörg Matthes, David 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
- Van Duyn, Emily, and Jessica 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
- Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146–1151. doi:10.1126/science.aap9559
- Vraga, Emily K., and Leticia 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
- Walter, Nathan, Jonathan Cohen, R. Lance Holbert, and Yasmin Morag. 2020. "Fact-Checking: A Meta-Analysis of What Works and for Whom." *Political Communication* 37 (3): 350–375. doi:10.1080/10584609.2019.1668894
- Walter, Nathan, and Sheila T. Murphy. 2018. "How to Unring the Bell: A Meta-Analytic Approach to Correction of Misinformation." *Communication Monographs* 85 (3): 423–441. doi:10.1080/03637751.2018.1467564
- Walter, Nathan, and Riva Tukachinsky. 2020. "A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It?" *Communication Research* 47 (2): 155–177. doi:10.1177/0093650219854600
- Wardle, Claire, and Hossein Derakhshan. 2017. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking..* Strasbourg: Council of Europe.
- Weingart, Peter, and Lars Guenther. 2016. "Science Communication and the Issue of Trust." *Journal of Science Communication* 15 (5): C01. doi:10.22323/2.15050301

Appendix

Table A1. Results of regression analyses where claim credibility is predicted by fact-check formats.

	(1) Partially false claim	(2) Accurate claim
Condition 0: Control condition: no fact-check	Ref.	Ref.
Condition 1: label [yes] & factuality [false]	-0.15 (0.13)	0.06 (0.11)
Condition 2: label [no] & factuality [false]	-0.02 (0.13)	-0.01 (0.11)
Condition 3: label [yes] & factuality [mostly false]	0.14 (0.13)	0.10 (0.11)
Condition 4: label [no] & factuality [mostly false]	0.02 (0.13)	-0.07 (0.11)
Condition 5: truth sandwich	0.14 (0.12)	-0.03 (0.11)
Constant	2.92*** (0.09)	3.91*** (0.08)
Observations	752	752

Note: Standard errors in parentheses. *** $p < 0.001$.

Table A2. Claim credibility predicted by whether the fact-check includes a label.

	(1) Partially false claim	(2) Accurate claim
Control condition: no fact-check	Ref.	Ref.
Fact-check label: yes	-0.00 (0.11)	0.08 (0.10)
Fact-check label: no	0.05 (0.10)	-0.03 (0.09)
Constant	2.92*** (0.09)	3.91*** (0.08)
Observations	752	752

Note: Standard errors in parentheses. *** $p < 0.001$.