



ELSEVIER

Contents lists available at ScienceDirect

## International Journal of Research in Marketing

journal homepage: [www.elsevier.com/locate/ijresmar](http://www.elsevier.com/locate/ijresmar)

# Familiarity backfire effects? Disentangling the competing effects of repetition and fact-checking corrections of brand misinformation

Ipek N. Nibat<sup>a</sup>, Olivier Trendel<sup>b</sup>, Robert Mai<sup>b,\*</sup>, Tinka Krüger<sup>c</sup>,  
Wassili Lasarov<sup>d</sup>, Stefan Hoffmann<sup>c</sup>

<sup>a</sup> Sabanci Business School, Department of Marketing, Sabanci University, Orta Mahalle, Universite Caddesi No:27 Tuzla, 34956 Istanbul, Turkey

<sup>b</sup> Grenoble Ecole de Management, Department of Marketing, 12 Rue Pierre Semard, 38000 Grenoble, France

<sup>c</sup> Kiel University, Institute of Business Administration, Department of Marketing, Westring 425, 24098 Kiel, Germany

<sup>d</sup> Audencia Business School, Department of Marketing, 8 Rte de la Jonelière, 44300 Nantes, France

## ARTICLE INFO

## Keywords:

Misinformation  
Illusory truth effect  
Correction  
Fact-checking  
Backfire effect

## ABSTRACT

Misinformation poses a growing threat to firms, distorting consumer beliefs and damaging brand evaluations. A common corrective strategy involves attaching fact-checking labels to false claims, yet concerns persist that such corrections may backfire by strengthening familiarity with the misinformation. Across five studies (N = 4337), this article systematically compares the competing effects of repetition and correction on belief in corporate misinformation and brand evaluations. Repetition reliably increases belief in misinformation (illusory truth effect), while correction typically offsets this effect and even reverses it with strong, unambiguous labels. This research finds no evidence of a familiarity backfire effect: in none of the studies, repetition increases belief in the misinformation more than correction reduces it. While brand evaluations are less affected by repetition, they do decline following exposure to misinformation and are only partially restored by corrections. The article further examines how brand familiarity and the timing of assessment shape these effects. Corrections are effective both immediately and after a delay, and benefit unfamiliar brands more than familiar ones. Finally, corrections issued at first exposure, reaching new audiences, also reduce belief in misinformation without backfiring during future exposures. These findings inform managerial decisions on misinformation response and contribute to understanding how misinformation familiarity and correction compete in shaping consumer judgments.

## 1. Introduction

Misinformation represents an existential danger for firms (Appel et al., 2020; Di Domenico et al., 2021; Sanchez et al., 2020). Examples of corporate misinformation include Walmart writing off its customers' donations, Converse replacing its all-star logo with a satanic symbol, or McDonald's banning the purchase of food for people without homes ("Business Archives," n.d.). To fight such misinformation, issuing a correction backed by independent fact-checkers is a common practice regularly implemented by organizations (Prike et al., 2023). Yet, because such a practice often includes the repetition, or at least, a reminder of the false claim, concerns

\* Corresponding author.

E-mail addresses: [ipek.nibat@sabanciuniv.edu](mailto:ipek.nibat@sabanciuniv.edu) (I.N. Nibat), [olivier.trendel@grenoble-em.com](mailto:olivier.trendel@grenoble-em.com) (O. Trendel), [robert.mai@grenoble-em.com](mailto:robert.mai@grenoble-em.com) (R. Mai), [marketing@bwl.uni-kiel.de](mailto:marketing@bwl.uni-kiel.de) (T. Krüger), [wlasarov@audencia.com](mailto:wlasarov@audencia.com) (W. Lasarov), [stefan.hoffmann@bwl.uni-kiel.de](mailto:stefan.hoffmann@bwl.uni-kiel.de) (S. Hoffmann).

<https://doi.org/10.1016/j.ijresmar.2026.03.007>

Received 12 December 2023; Received in revised form 18 March 2026; Accepted 26 March 2026

Available online 27 March 2026

0167-8116/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Please cite this article as: Ipek N. Nibat et al., *International Journal of Research in Marketing*, <https://doi.org/10.1016/j.ijresmar.2026.03.007>

have been raised that it might boost familiarity with the claim and, ironically, enhance perceived trustworthiness and reliance on the very misinformation that is being corrected, leading to a “*familiarity backfire effect*.” While some studies found no evidence of such a backfire effect (Ecker et al., 2023; Swire et al., 2017) or even found that the familiarity effect actually increases the effectiveness of corrections (Ecker et al., 2017, 2019; Kemp et al., 2022; Wahlheim et al., 2020), several studies demonstrated a greater belief in misinformation or more use of misinformation following a correction due to an increase in familiarity (Berinsky, 2017; Pluviano et al., 2017, 2019; Skurnik et al., 2005, 2007; Thomas & Autry, 2024). Notably, familiarity backfire effects appear more likely in brief corrections to misinformation (Thomas & Autry, 2024), such as when using fact-checking by attaching labels to claims based on the evaluations of independent fact-checkers that rate claims by their truthfulness (e.g., Politifact, Snopes). Yet, evidence for a familiarity backfire effect remains mixed with a recent call to “more specifically investigate how familiarity boosts that increase belief and corrections that decrease belief interact” (Swire-Thompson et al., 2020).

This research addresses this call by relying on an experimental design that enables to directly compare the competing effects of 1) correction with a fact-checking label and 2) familiarity due to repetition. Evidence of a familiarity backfire effect is obtained if the effect of repetition is larger than the effect of correction. It is well established that people tend to judge the information they have seen before as more accurate, regardless of its veracity. This phenomenon, called the “*illusory truth effect*” (Arkes et al., 1989; Begg et al., 1992; Hasher et al., 1977; Pennycook et al., 2018), is pervasive and difficult to prevent (Schwarz et al., 2021). While correction with fact-checking was found ineffective in mitigating the effect of repetition (Pennycook et al., 2018), the relative size of the effects of repetition and correction with fact-checking labels has not been assessed so far. Interestingly, the results of Pennycook et al. (2018, cf. supplementary material for Studies 2 and 3) suggest that the effect of repetition can be larger than the effect of correction with fact-checking, but the authors did not provide a formal test of this comparison. As a first goal, this research aims to assess the possibility of a familiarity backfire effect by systematically disentangling and comparing the competing forces of repetition and correction on belief in misinformation.

To confirm the managerial relevance of this issue, we conducted a pilot study with 301 full-time managers and senior managers (Web Appendix A). Most managers (69%) heard about a possible backfire effect when issuing corrections to fight misinformation and they expressed a substantial degree of concern about it (67%, where 0% = “not at all concerned” and 100% = “very concerned”). Interestingly, managers who encounter misinformation more frequently in their daily lives expressed stronger concerns about a backfire effect. Managers thus widely recognize the potential backfire effect of misinformation corrections and are generally worried about it, which might hinder the likelihood of firms issuing corrections.

This is a dilemma for firms as people can continue to rely on misinformation in their reasoning even after being given corrections and even when the corrections substantially reduced belief in misinformation (Lewandowsky et al., 2012; Tybout et al., 1981). It is thus essential to also assess the possibility of a familiarity backfire effect beyond belief in misinformation by considering continued reliance on the misinformation. Notably, reduced belief in misinformation may not necessarily translate into evaluations of the target of the misinformation (Jussim & Zanna, 2005). Recent findings in the political domain suggest that the effect of repetition on belief in misinformation and evaluations of the target differ (Mattavelli et al., 2024). Accordingly, as a second goal of this research, we investigate the competing effects of repetition and correction with fact-checking labels on a relevant marketing outcome for managers, the evaluations of brands targeted by misinformation.

Across five studies (N = 4337), we assess the possibility of a familiarity backfire effect and decompose the mechanisms of repetition and correction with fact-checking labels on consumers’ belief in corporate misinformation and evaluations of the targeted brands. In Study 1 to 3, we comprehensively search for a backfire effect by considering different types of correction labels varying in strength, different levels of brand familiarity (i.e., well-known versus unknown brands), and also different timing of assessment (i.e., immediately after misinformation correction versus at a later stage). In Study 4 and its follow-up study, we assess the potential backfire effect of misinformation correction when it reaches a new audience; that is, when the correction reaches the consumer during the first exposure to the misinformation (compared to during the second exposure for studies 1 to 3). Because a correction tends to increase processing of misinformation and, hence, potentially increases the repetition effect (Autry & Duarte, 2021; Prike et al., 2023), it is unclear if such an initial corrected exposure could backfire compared to an uncorrected initial exposure.

## 2. Theoretical background

### 2.1. Firms victims of misinformation

Misinformation features claims that look like genuine news without any factual basis. Unlike rumors and gossip, misinformation is factually false, and its falsehood is verifiable. It also differs from deceptive advertising that promotes brands (e.g., Kopalle & Lehmann, 2015; Rao, 2022), as misinformation appears with denigrating content (Thorson, 2016). Firms have a paradoxical relationship with misinformation. On one hand, they may advertise on and benefit from high-traffic misinformation platforms. On the other hand, misinformation can target a firm and hurt it by biasing the opinions of consumers, leading to boycotts or hurting brand reputations (Berthon & Pitt, 2018; Di Domenico et al., 2021). The present research focuses on the second aspect: misinformation targeting firms.

Examining misinformation effects on firms is crucial. In 2017, major financial entities such as S&P and Ethereum reported losses of \$341 million and \$4 billion, respectively, due to misinformation (Cavazos, 2019). While one might expect consumer perceptions of familiar brands to be resilient, evidence shows that even global brands like Pepsi, New Balance, United Airlines, or Starbucks are vulnerable, suffering losses in profits, stock value, and reputation from misinformation (Sanchez et al., 2020). Despite this growing threat, research remains limited, prompting increasing calls for more comprehensive studies, especially in marketing (e.g., Appel et al., 2020; Di Domenico et al., 2021; Johar, 2022).

## 2.2. Correcting misinformation: The role of fact-checking

While policymakers are concerned with generating general resistance against misinformation to limit the threat to society at large, managers are predominantly concerned about mitigating the imprint of a particular piece of misinformation specifically targeting their business. Given these different foci, a news-specific investigation of corporate misinformation is needed. A popular news-specific tactic is attaching labels to claims based on evaluations by independent fact-checkers (e.g., Politifact, Snopes) that rate claims by their truthfulness. Even though fact-checks help people make better, evidence-based decisions, the optimal method of fact-checking remains unclear (Amazeen et al., 2018; Ecker et al., 2020). Politifact uses a “Truth-o-meter”, a colorful, visual indicator of a claim’s truthfulness (Holan, 2018). Social media platforms used to flag misinformation as “disputed” (Mosseri, 2017). Beyond its form, the fact-checker’s credibility is crucial for its effectiveness (Ecker et al., 2020, 2022). In fact-checking, credibility is often established by the presence of neutral third parties who do not benefit from disproving a claim.

Overall, there is little clarity on whether fact-checking works effectively and under which circumstances (Mende et al., 2023). Warnings were found to have limited efficacy about a week after exposure (Ecker et al., 2020) and were found ineffective for older people in the long term (Skurnik et al., 2005). More worrying, tagging misinformation and recirculating it for the sake of correction may backfire by creating a familiarity effect (Skurnik et al., 2005).

## 2.3. The familiarity backfire effect

The backfire effect occurs when a correction inadvertently increases belief in the very misconception it seeks to refute, and it is often cited as a reason for avoiding corrections altogether (Swire-Thompson et al., 2020). Two types of backfire effects have been identified in the literature: the worldview backfire effect and the familiarity backfire effect. Both result in increased belief after a correction, but they stem from different psychological mechanisms. The worldview backfire effect arises when a correction challenges an individual’s deeply held beliefs, leading to motivated reasoning and greater scrutiny of counter-attitudinal information (Cook & Lewandowsky, 2012). In contrast, the familiarity backfire effect, the focus of this research, occurs when misinformation is repeated within a correction (Schwarz et al., 2007). Repetition increases the perceived accuracy of information, a phenomenon known as the illusory truth effect (Arkes et al., 1989; Begg et al., 1992; Hasher et al., 1977; Pennycook et al., 2018). This effect is robust and primarily explained by processing fluency. The ease with which information is processed fosters a sense of familiarity and, consequently, perceived truthfulness (Greifeneder et al., 2021; Walter & Tukachinsky, 2020). When individuals encounter the same misinformation repeatedly, even in the context of a correction, familiarity can override memory for the corrective details, leading to reliance on the misinformation (Schwarz et al., 2016). Thus, while the illusory truth effect occurs in the absence of corrective information, and the familiarity backfire effect in its presence, both are thought to share the same underlying mechanism of processing fluency (Unkelbach et al., 2019).

Whether a correction backfires depends on the interplay between two cognitive processes: automatic familiarity and strategic processing (Chaiken & Maheswaran, 1994; Swire et al., 2017). Familiarity is a fast, effortless process that signals whether information has been encountered before, whereas strategic processing involves deliberate evaluation of available cues (such as a corrective label) to determine accuracy. When strategic processing is weak or underutilized, judgments may default to familiarity, which can lead to the persistence of misinformation even when corrective information is present (Ecker et al., 2010).

Importantly, this competition can unfold in two ways, depending on cues available at the moment of judgment. When a correction cue is present, strategic processing can operate at the time of judgment by using the label as a diagnostic input. When the cue is absent, judgments depend more on memory processes, specifically, what is retrieved from prior exposure (e.g., a misinformation trace, a correction trace, or their relative accessibility). Thus, the risk of a backfire effect is highest when familiarity is strong and either (a) diagnostic correction cues are not used or (b) corrective information is not retrieved.

This distinction between cue-based evaluation and memory-based retrieval is reflected in prior work. The seminal work on which the familiarity backfire effect builds is Skurnik et al. (2005), noteworthy because it was published in a marketing journal. They found that when the “false” tags were no longer present at the time of assessment, participants became more likely over time to misremember previously tagged health-related claims as true, consistent with contextual details fading faster than familiarity.

Although theory suggests that familiarity can undermine corrections, empirical findings are mixed (Table 1). Some studies report no evidence of a familiarity backfire effect (e.g., Ecker et al., 2019, 2023; Swire et al., 2017) and even show repetition can enhance correction effectiveness (Ecker et al., 2017; Kemp et al., 2022; Wahlheim et al., 2020). Other work finds that corrections increase belief in misinformation or reliance on it, especially when brief and minimal, as with fact-checking labels (e.g., Berinsky, 2017; Peter & Koch, 2016; Pluviano et al., 2017, 2019; Skurnik et al., 2005). Backfire also seems more likely when corrections provide little contextual justification, which is often true for fact-checking labels (Thomas & Autry, 2024). Overall, evidence remains inconsistent, prompting calls to clarify how familiarity and correction compete (Swire-Thompson et al., 2020). Addressing this gap requires comparing the relative strength of repetition and correction, central to this research.

## 2.4. Factors affecting the familiarity backfire effect

The relative influence of automatic familiarity and strategic processing depends on contextual factors. When corrections are detailed and explicit, they provide strong cues that facilitate strategic evaluation, reducing reliance on familiarity (Prike & Ecker, 2023). When they are minimal or ambiguous, or when individuals fail to engage with the corrective cue, familiarity may dominate (Dechêne et al., 2010). Higher cognitive load or lower motivation can similarly shift judgments toward automatic processing. These

**Table 1**  
Overview of the Familiarity Backfire Effect Literature.

Article	Misinformation topic	DV		Assessment of the differential effect of...			Did backfire occur?
		Belief / Memory	Downstream consequence	...correction type	...target familiarity	...timing of assessment	
Skurnik et al., 2005	Health	✓		Brief correction		30 min. or 3 days later	✓
Skurnik et al., 2007 <sup>a</sup>	Health / Vaccines	✓		Brief correction		Immediate or 30 min. later	✓
Wilson & Park, 2008	Health	✓		Simple negation		45-min. delay only	✓
Cameron et al., 2013	Health / Vaccines	✓		Detailed correction with/without justification		Up to one week only	
Peter & Koch, 2016	Health	✓	✓	Brief correction		A few min. & 5 days later	✓
Rich & Zaragoza, 2016	Warehouse fire	✓		Detailed correction	Only unfamiliar	Immediate assessment only	
Berinsky, 2017	Death panels	✓	✓	Non-partisan vs. republican vs. democratic correction		Immediate & one month later	Mixed
Ecker et al., 2017	Warehouse fire	✓		Retraction with subtle / explicit / no reminder		30 min. delay only	
Pluviano et al., 2017	Health / Vaccines	✓	✓	Myth vs. fact/ visual / fear correction		Immediate or 7 days later	✓
Swire et al., 2017	Various	✓		Brief or detailed correction		Immediate, 30 min, one or three weeks later	
Ecker et al., 2019	Various	✓		Brief or detailed correction		One day later only	
Pluviano et al., 2019	Health / Vaccines	✓	✓	Myth vs. fact correction		Immediate or 7 days later	Mixed
Carnahan & Garrett, 2020	Health / GMO	✓		Myth vs. fact correction	Measured	Online- vs. memory-based	
Ecker et al., 2023	Health / Vaccines	✓	✓	Myth vs. fact / visual / fear correction		Immediate or one week later	
Thomas & Autry, 2024	Various	✓		Licensed vs. unlicensed correction	Only unfamiliar	Immediate assessment only	✓
<b>Current Research</b>	<b>Marketing</b>	✓	✓	<b>Ambiguous vs. unambiguous correction</b>	<b>Well-known vs. unknown brands</b>	<b>Immediate with corrective cues present vs. delayed without cue</b>	

Notes. The extended version of this table is available in [Web Appendix B](#). Even though some papers listed (e.g., Skurnik et al., 2005; Peter & Koch, 2016) did not include a no-correction baseline condition, we still include them in our literature review, as the familiarity backfire effect term has also been used for myths that are misremembered as facts. We only include papers that focus on the familiarity backfire effect and exclude those that focus on the worldview backfire effect. <sup>a</sup> unpublished manuscript.

insights suggest that the effectiveness of corrections and the risk of familiarity-driven errors depend on how corrections are presented, the characteristics of the information, and the timing of evaluation.

Building on this framework, the present research examines three factors that could shape the competing impacts of repetition and correction: (1) the type of fact-checking label, (2) the familiarity of the brand targeted by misinformation, and (3) the timing of the assessment. These factors are theoretically relevant because they shift the balance between familiarity-based responding and correction-based responding by changing (i) the diagnosticity and use of corrective cues at judgment and/or (ii) reliance on memory retrieval when these cues are delayed or absent. We next discuss each factor and why its influence remains unclear.

#### 2.4.1. Type of fact-checking label

Fact-checking labels differ in how clearly they communicate that a claim is false. In this research, we compare two existing formats across all studies (except Study 3): one that unambiguously tags the misinformation as “False based on 3rd party fact-checkers” and another that uses more cautious language, “Disputed by 3rd party fact-checkers” (Mosseri, 2017). While we use these labels as a manipulation of corrective clarity, they may also differ in implied meaning: “false” conveys a categorical verdict, whereas “disputed” can signal contestation or interpretive uncertainty, providing weaker grounds for confident inference at judgment.<sup>1</sup> A similar “disputed” label has been used on social media for several years, and its efficacy has been questioned (e.g., Clayton et al., 2020; Pennycook et al., 2018). From a dual-process perspective, clear labels provide stronger cues for strategic processing, making it easier for individuals to override the automatic sense of familiarity associated with the repeated misinformation. Ambiguous labels, by contrast, may leave room for doubt and require more effortful interpretation. This could either reduce reliance on familiarity, if individuals engage in deeper processing, or increase it when motivation or cognitive resources are low, leading people to default to familiarity (Chan et al., 2017; Hart et al., 2009). Detailed corrections enhance belief updating (Swire et al., 2017), but labels are inherently brief and lack explanatory content. It is therefore unclear whether stronger wording alone is sufficient to counteract familiarity or whether both types of labels remain vulnerable to its influence.

#### 2.4.2. Brand familiarity

The familiarity of the brand targeted by misinformation may shift the relative weight of automatic and strategic processes. Brand familiarity can plausibly affect repetition effects in either direction. Compared with well-known brands, unfamiliar brands provide fewer prior anchors, which could increase reliance on repetition-based fluency when judging accuracy. However, well-known brands also provide richer schemas (including unfavorable ones) that can make allegations feel more diagnostic and plausible, potentially strengthening the extent to which repetition-induced fluency is interpreted as truth. By contrast, claims about unfamiliar brands may be seen as less diagnostic because consumers lack background knowledge to interpret them, which can limit how much repetition increases belief. At the same time, corrections may still increase the accessibility of the claim content for unfamiliar brands, as the allegation may be among the few brand-related cues encoded (Ecker et al., 2020). To address these competing possibilities, we compare well-known and unknown brands in Study 2.

#### 2.4.3. Timing of assessment

The timing of assessment can influence whether consumers rely more on strategic processing or automatic familiarity. Familiarity is relatively stable over time, whereas the ability to strategically apply corrective information can diminish as attention or motivation wanes (Knowlton & Squire, 1995). Swire et al. (2017) observed that corrections were effective immediately but weakened after a delay, particularly for myths, which showed partial rebound. This suggests that delays increase the likelihood of judgments being driven by familiarity rather than deliberate evaluations of corrective cues. In our context, this issue is especially relevant because, in the case of introducing a delay (tested in Study 3), consumers are re-exposed to the misinformation without the corrective label at the time of assessment, and therefore have to rely on their memory. Decision-making under such conditions could amplify the influence of familiarity, and thus repetition effects, making it an open question whether delayed evaluation under these conditions reduces or reverses the benefits of prior correction.

### 2.5. Misinformation’s continued influence and downstream effects

Misinformation can exert a continued influence on judgments even after being corrected, a phenomenon widely documented in cognitive psychology (Ecker et al., 2022; Johnson & Seifert, 1994; Lewandowsky et al., 2012; Walter & Tukachinsky, 2020). Corrections can reduce belief in misinformation but rarely eliminate its impact entirely, as people may still rely on the original claim when making decisions. A classic example is that even when people disbelieve the claim “McDonald’s meat contains worms,” they still evaluated eating at McDonald’s less positively (Tybout et al., 1981). This persistence is particularly problematic in marketing, where believing misinformation can inhibit purchases (Fong et al., 2024).

While most research has focused on accuracy perceptions, downstream consequences such as brand evaluations remain under-explored. Yet, these outcomes are highly relevant for managers because brand evaluations inform purchase intentions and loyalty. Repeated exposure to brand-denigrating misinformation can increase processing fluency, which can shape evaluations—positively

<sup>1</sup> We do not claim that “false” and “disputed” differ only in ambiguity; rather, ambiguity is the focal dimension for our theorizing. Pretests 2 and 3 reported in the [Web Appendices D and E](#) show that both labels are perceived similar in signaling low accuracy/untruthfulness and receive comparable visual attention, consistent with our comparison, while still differing in perceived clarity/certainty.

when associated with favorable cues, but negatively when linked to harmful content (Corneille et al., 2020; Lee & Labroo, 2004). Moreover, fluency may indirectly affect evaluations by increasing perceived accuracy, which in turn guides evaluations (Fishbein & Middlestadt, 1995). However, the link between believing a claim and evaluating a brand does not imply that experimentally shifting belief in a claim through repetition or correction will necessarily translate into detectable changes in brand evaluations, which may be more stable and anchored in prior attitudes. For example, repeated exposure to political misinformation increased perceived accuracy without altering attitudes toward the candidate (Mattavelli et al., 2024), suggesting that evaluations may be less elastic than belief updating. Hence, it is essential to examine whether, and under what conditions, repetition and correction affect not only belief in misinformation but also brand evaluations. Addressing this gap, the present research investigates the competing impacts of repetition and fact-checking corrections on both belief in misinformation and consumers' evaluations of the targeted brands.

### 3. Overview of main studies

We conducted a series of studies, including a pilot study (Web Appendix A), three pretests (Web Appendices C/D/E), and five main studies (Table 2). As brands targeted by misinformation campaigns often tend to be major brands, we used actual misinformation about firms containing verifiably false, misleading information (commonly known as “fake news”) in all our studies (except in Study 2, in which we systematically manipulate brand familiarity, using misinformation about fictitious brands for half of the brands). To ensure ecological validity, we gathered “fake news” but also true news from Snopes.com, an established fact-checking platform (Pennycook et al., 2021). To determine the different types of misinformation effects, our studies adopted a validated multistage approach (Pennycook et al., 2018): a familiarization stage with a set of factually false and true news presented in random order, a distractor stage for temporal separation (Podsakoff et al., 2003), and an assessment stage with the news from Stage 1 (which should be familiar) and additional news stories (which are new). In this stage, participants reported their beliefs for each story and brand evaluations.

This design allows us to differentiate the impact of repeated versus novel misinformation exposures, capturing the *repetition effect* for both belief in misinformation and brand evaluations. For the misinformation *correction effect*, we used fact-checking labels, manipulating their strength (“false” or “disputed”), and observed their impact on belief and brand evaluations compared to the condition in which misinformation is not corrected. When fact-checking labels are used to correct circulating misinformation, their ability to reduce the impact of the misinformation (the *correction effect*) competes with the tendency of repeated information to appear more believable (the *repetition effect*). As a result, the overall effect of using fact-checking (net effect) is less than the label's correction effect alone. We obtain evidence for a *familiarity backfire effect* if the negative effect of repetition exceeds the positive effect of correction. For parsimony, we report the most relevant tests in the main text; the Web Appendices F, G, and H contain all the statistical details, manipulation checks (all were successful), and experimental material. Note that some preregistered analyses are reported in the Web Appendix, as they are not central to the revised focus of this article.

**Table 2**  
Overview of Study Objectives and Key Findings.

Study	Objective	Key Findings
<b>Pilot Study (WA)</b> N = 301	Managerial relevance of the backfire effect	Managers are knowledgeable and worry about the correction of corporate misinformation provoking a backfire effect.
<b>Pretests 1 to 3 (WA)</b> Total N = 462	Validation of material	Experimental materials (i.e., true and false news stories, correction labels) are validated.
<b>Study 1</b> N = 466	Backfire effect for well-known brands; assessing “false” and “disputed” labels	Repetition increases belief in misinformation. Strong labels (“false”) reverses this effect, reducing belief beyond repetition, while weaker labels (“disputed”) only offset repetition without added benefit. No evidence of familiarity backfire. Brand evaluations are unaffected by repetition or correction.
<b>Study 2</b> N = 987	Backfire effect for well-known vs. unknown brands; no-news baseline to gauge overall damage	Repetition increases misinformation belief for both brand types. Strong labels reverse this effect for well-known and unknown brands, while weaker labels only offset repetition. Brand evaluations are unaffected by repetition; strong labels improve evaluations for unknown brands but do not restore them to the no-news baseline. No evidence of familiarity backfire.
<b>Study 3</b> N = 490	Assess whether backfire emerges when assessment occurs later without label present (memory-based retrieval)	Repetition increases belief in misinformation. When assessment occurs later without the correction label present, strong labels (“false”) still reverse this effect, reducing belief beyond repetition, while brand evaluations improve modestly but remained low overall. No evidence of familiarity backfire.
<b>Study 4</b> N = 1400	Backfire effect for new audiences (label at first exposure, unlabeled at repeated exposure)	Repetition increases belief in misinformation. When new audiences first encounter misinformation with a correction label present, strong labels (“false”) reverse this effect on belief in misinformation, while weaker labels (“disputed”) only offset repetition. Brand evaluations are unaffected by repetition and show an improvement only with strong labels. No evidence of familiarity backfire.
<b>Follow-up to Study 4 (WA)</b> N = 994	Examine persistence of effects after a three-day delay	After a three-day delay, both labels (“false” and “disputed”) reverse the effect of repetition on belief in misinformation. Brand evaluations are unaffected by repetition and both labels only offset the effect of repetition. No evidence of familiarity backfire.

#### 4. Study 1: Searching for backfire effects for well-known brands

The goal of Study 1 is to investigate whether correcting misinformation by recirculating it with fact-checking labels attached can actually backfire. Does the correction of misinformation trigger recipients to discount the misinformation, or does the repeated exposure to the news even strengthen recipients' belief in the misinformation? In Study 1, we disentangle, quantify, and compare the effect of (i) correcting misinformation and thereby (ii) repeating it on (iii) consumers' beliefs in the misinformation and on their evaluations of the targeted brands. We also compare the effectiveness of two fact-checking labels varying in strength ("false" vs. "disputed"). Study 1 is preregistered: <https://doi.org/10.17605/OSF.IO/58FGV>.

##### 4.1. Procedure

Study 1 features a  $3 \times 2 \times 2$  mixed design, with correction label (no-correction vs. "disputed" label vs. "false" label) manipulated between-subjects, and exposure (repeated vs. novel) and news type (true vs. false) manipulated within-subjects. The study design visualized in Fig. 1 adopts a protocol established in prior research (Pennycook et al., 2018). In the first (familiarization) stage, participants saw 6 factually false and 6 factually true negative stories about well-known firms in random order, none of them carrying any correction label. To ensure engagement with each story, participants indicated their readiness to share these stories on social media (options: Yes/Maybe/No). In the second (distractor) stage, designed to separate the first and third stages temporally and psychologically (Podsakoff et al., 2003), we collected demographic information, and asked participants to search for patterns in five unstructured Jackson Pollock paintings. This task was unrelated to our study. Finally, in the assessment stage, participants saw the 12 stories from the first stage, and 12 additional novel stories (6 false, 6 true). In this stage, the 12 false stories were presented with either no-correction (control), or a "false" label ("false based on 3rd party fact-checkers"), or a "disputed" label ("disputed by 3rd party fact-checkers"). The true news stories did not carry any label. In this third stage, the 24 stories denigrating well-known firms (12 false, 12 true) appeared in random order, half of the stories being familiar to the participants (*repeated*), whereas the other half appeared for the first time (*novel*). To avoid news-specific distortions or order effects, one out of every two true stories and one out of every two false stories were randomly presented in stage 1. Also, in stage 3, all true and false stories were presented in random order.

##### 4.2. Materials and measures

Each story featured harmful content about a particular, well-known brand (e.g., "Levi's jeans' famous "two horse" logo represents an enslaved person is being killed," "P&G profits were donated to the Church of Satan," "Workers at the Corona brewery urinate into the beer"), representing various domains (e.g., consumer goods, digital platforms, fashion, retailing) to ensure the findings' generalizability to various industries. In the assessment stage, following Pennycook et al. (2018), participants indicated their familiarity with each news story, as a manipulation check ("Have you seen this news before in this study?", 0 = "No/Unsure," 1 = "Yes"). Then

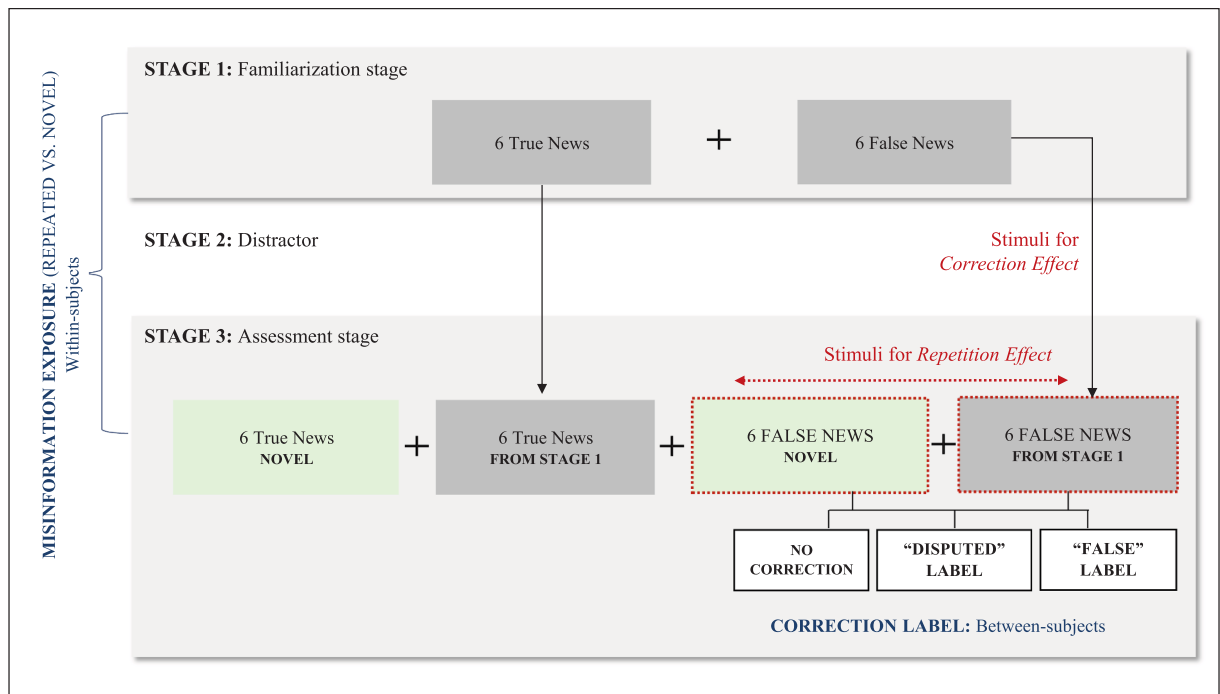


Fig. 1. Experimental Design, Study 1.

they reported their belief (“To the best of your knowledge, how accurate is the claim in the above headline?”) on a four-point scale (Pennycook et al., 2018; 1 = “Not at all accurate,” 4 = “Very accurate”) and evaluated the brand (“How do you feel towards the brand mentioned in the above news?,” 1 = “Negative,” and 7 = “Positive”). We generated index scores of these variables for all repeated and novel stories. Finally, we included a few preregistered exploratory measures that had no effect on our findings (e.g., animosity towards big brands) and provided a debrief to the participants.

#### 4.3. Sample

Study 1 was administered to 509 U.S.-based Prolific workers. After applying the preregistered exclusion criteria that filter the participants based on random responses, search for headlines on the internet, and failure to read the content (Pennycook et al., 2018), the final sample consisted of 466 participants ( $M = 28.18$  years,  $SD = 10.02$ , 82% women). The detailed sample exclusion protocol is provided in [Web Appendix I](#). Note that the same results hold when the preregistered exclusion criteria are not applied.

#### 4.4. Results

**Modeling framework.** We estimated two repeated-measures ANOVA with belief in misinformation and brand evaluations as the dependent variables, the news type (true vs. false, within-subjects), the exposure (repeated vs. novel, within-subjects), and the correction label (no-correction vs. “disputed” label vs. “false” label, between-subjects) as independent variables. For *belief in misinformation*, all main effects of all factors are significant (all  $ps < 0.001$ ), but the three-way interaction is not significant ( $F(2, 463) = 2.89, p = 0.057$ ). A two-way interaction arises between the correction label and news type ( $p < 0.001$ ); the other two-way interactions are non-significant (all  $ps \geq 0.229$ ). For *brand evaluations*, the RM-ANOVA shows a significant main effect for news type ( $F(1, 463) = 19.38, p < 0.001$ ) and a two-way interaction between the correction label and exposure ( $F(2, 463) = 4.28, p = 0.014$ ). The three-way interaction ( $F(2, 463) = 1.73, p = 0.179$ ) as well as other main effects or two-way interactions are non-significant (all  $ps \geq 0.136$ ).

We next focus on the (theoretically) relevant contrasts to disentangle the effect of repetition and correction. In a first step, we examine the belief in misinformation and determine the extent of the repetition effect. In a second step, we investigate the extent to which these implications can be prevented by correcting the misinformation. We then repeat these analyses for brand evaluations. [Table 3](#) presents an overview of the results of the distinct effect types tested. To quantify the repetition and correction effects, we specify two different baselines. To determine the repetition effect ( $\Delta_{\text{repetition}}$ ), our experimental design systematically includes misinformation that was seen repeatedly versus misinformation seen only once. This *within-subjects* factor allows us to quantify the consequences of repetition for belief in the misinformation (“illusory truth”) and brand evaluation. To assess whether brand-related misinformation can be effectively corrected when repeating it, we determine the impact of the correction label ( $\Delta_{\text{correction}}$ ). To this end, the *between-subjects* factor of our experiment includes a control condition in which none of the news stories are corrected, serving as the baseline to quantify the effect of correcting with a strong (“false”) or a somewhat weaker (“disputed”) correction label. This allows us to determine whether the detrimental effect of repetition is larger, equal, or smaller than the correction effect ( $\Delta_{\text{net effect}}$ ).

**Repetition effect for belief in misinformation.** We find a repetition effect for belief in misinformation ([Table 3](#)), confirming that participants fall victim to the repetition, also known as the “illusory truth” effect ( $M_{\text{repeated}} = 2.40, M_{\text{novel}} = 2.24, \Delta_{\text{repetition}} = 0.15, p < 0.001, d = 0.28$ ). That is, participants judge the repeated misinformation more accurate than the misinformation they have seen only once ([Table 4](#)).

**Correction effect for belief in misinformation.** To determine corrections’ impact, we estimate the contrast of the corrected misinformation compared to the no-correction control condition. The correction of brand-related misinformation leaves measurable traces on consumers’ belief in misinformation, confirming the effectiveness of the “false” label, which significantly reduces consumers’ belief ( $M_{\text{no-correction}} = 2.40$  vs.  $M_{\text{false}} = 1.81, \Delta_{\text{correction}} = -0.59, p < 0.001, d = 1.10$ ). The weaker “disputed” label also reduces belief in misinformation ( $M_{\text{disputed}} = 2.22, \Delta_{\text{correction}} = -0.18, p = 0.011, d = 0.33$ ) but less so than the “false” label.

**Net effect for consumers’ belief in misinformation.** Contrasting the repetition and correction effects, we do not obtain evidence that correction backfires by repeating the very piece of misinformation. On the contrary, the correction effect of the “false” label is significantly greater than the misinformation repetition effect ( $\Delta_{\text{net effect}} = -0.43, p < 0.001, d = 0.81$ ), implying that this elicits a reversal ([Table 3](#)): the correction of one’s belief extends beyond merely compensating for the illusory truth effect. Yet, the “disputed” label does not elicit such a reversal net effect, although it does not leave a repetition-induced backfire effect either. Comparing the repetition effect vs. correction with “disputed” label reveals that this only offsets the detrimental effect of repetition ( $\Delta_{\text{net effect}} = -0.02, p = 0.723, d = 0.04$ ) but does not have additional benefits for reducing one’s belief in the misinformation.

**Repetition effect for brand evaluations.** As a second managerially relevant outcome, we shift the focus to brand evaluations, which are correlated to consumers’ beliefs in misinformation denigrating the target brands ( $r = -0.233, p < 0.001$ ). Despite these beliefs being inflated due to repetition, we do not observe a significant repetition effect for brand evaluations ( $M_{\text{repeated}} = 3.66, M_{\text{novel}} = 3.63, \Delta_{\text{repetition}} = 0.04, p = 0.575, d = 0.04$ ).

**Correction effect for brand evaluations.** For the correction effect, we only observe a tendency that correction of misinformation can help prevent a drop in brand evaluation, but this correction effect does not reach significance for the “false” fact-checking label ( $M_{\text{no-correction}} = 3.66$  vs.  $M_{\text{false}} = 3.79, \Delta_{\text{correction}} = 0.13, p = 0.183, d = 0.14$ ). The same is true for the “disputed” label ( $M_{\text{disputed}} = 3.63, \Delta_{\text{correction}} = -0.03, p = 0.762, d = 0.03$ ).

**Net effect for brand evaluations.** Considering the lack of significant repetition and correction effects, we do not observe a backfire effect. For both labels, the resulting net effect is not significant (“false” label:  $\Delta_{\text{net effect}} = 0.17, p = 0.097, d = 0.18$ ; “disputed” label:  $\Delta_{\text{net effect}} = 0.01, p = 0.958, d = 0.01$ ).

**Table 3**  
Overview of the Results of all Studies.

	Belief in Misinformation				Brand Evaluations					
	Repetition Effect	Correction Effect		Net Effect ( $\Delta$ )		Repetition Effect	Correction Effect		Net Effect ( $\Delta$ )	
		Disputed	False	Disputed	False		Disputed	False		
<b>Study 1</b>	0.15***	-0.18*	-0.59***	-0.02 <sup>ns</sup> (O)	-0.43*** (R)	0.04 <sup>ns</sup>	-0.03 <sup>ns</sup>	0.13 <sup>ns</sup>	0.01 <sup>ns</sup> (O)	0.17 <sup>ns</sup> (O)
<b>Study 2</b>										
Well-known brands	0.30***	-0.21***	-0.42***	0.09 <sup>ns</sup> (O)	-0.12* (R)	0.04 <sup>ns</sup>	0.01 <sup>ns</sup>	0.01 <sup>ns</sup>	0.05 <sup>ns</sup> (O)	0.05 <sup>ns</sup> (O)
Unknown brands	0.23***	-0.14*	-0.44***	0.09 <sup>ns</sup> (O)	-0.21*** (R)	0.05 <sup>ns</sup>	-0.03 <sup>ns</sup>	0.27*	0.02 <sup>ns</sup> (O)	0.32** (R)
<b>Study 3</b>	0.18***	-	-0.48***	-	-0.29*** (R)	-0.05 <sup>ns</sup>	-	0.19*	-	0.14 <sup>ns</sup> (O)
<b>Study 4</b>	0.21***	-0.26***	-0.54***	-0.05 <sup>ns</sup> (O)	-0.33*** (R)	-0.07 <sup>ns</sup>	0.04 <sup>ns</sup>	0.22**	-0.02 <sup>ns</sup> (O)	0.16* (R)

Notes. Repetition effect ( $\Delta$  repeated vs. novel): mean difference between the repeated and novel news in the no-correction condition. Correction effect ( $\Delta$  label vs. no-correction): mean difference of the repeated news between the no-correction and the correction label conditions. Levels of significance: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , ns: not significant ( $p > 0.05$ ). Type of net effect: (R) Reversal (correction effect > repetition effect), (O) Offset (correction effect = repetition effect), (B) Backfire (repetition effect > correction effect).

**Table 4**  
Descriptive Statistics of all Studies.

	Belief in Misinformation						
		Novel			Repeated		
		No-Correction	Disputed	False	No-Correction	Disputed	False
Study 1	2.24	2.07	1.70	2.40	2.22	1.81	
Study 2							
Well-known brands	2.21	2.07	1.94	2.51	2.30	2.09	
Unknown brands	2.28	2.15	1.94	2.51	2.37	2.07	
Study 3	2.14	-	2.05	2.33	-	1.85	
Study 4	2.10	2.03	2.02	2.31	2.05	1.77	
	Brand Evaluations						
		Novel			Repeated		
		No-Correction	Disputed	False	No-Correction	Disputed	False
Study 1	3.63	3.61	3.86	3.66	3.63	3.79	
Study 2							
Well-known brands	3.83	3.68	3.76	3.87	3.88	3.88	
Unknown brands	3.29	3.39	3.43	3.33	3.30	3.60	
Study 3	3.59	-	3.69	3.54	-	3.73	
Study 4	3.68	3.70	3.75	3.61	3.65	3.83	

Notes. Values represent means. Reported values refer only to the misinformation condition (news type = false). A complete set of means and standard deviations, including those for true stories, is provided in [Web Appendix F](#).

#### 4.5. Discussion

Study 1 examined whether correcting misinformation about well-known brands can backfire. Results show a clear repetition effect on consumers' belief in the misinformation: repeated exposure increases perceived accuracy, consistent with the illusory truth effect. However, this effect did not extend to brand evaluations, suggesting that attitudes toward established brands are relatively resilient, even when misinformation is repeated.

Corrections reduced belief in corporate misinformation, particularly when using the stronger "false" label, thereby more than offsetting the repetition effect. The weaker "disputed" label only neutralized the repetition effect without producing additional benefits. These findings indicate that when misinformation is repeated, the corrective cue must be strong enough to reverse familiarity-driven processing. Importantly, we find no evidence of a familiarity backfire effect; the repetition effect was never larger than the correction effect.

Finally, neither repetition nor correction significantly influenced brand evaluations, despite their correlation with belief in the misinformation about the brand. This suggests that while misinformation can distort accuracy judgments, its immediate imprint on attitudes toward well-known brands is limited. From a managerial perspective, clear corrections are effective for belief, and concerns about backfire appear unwarranted in this context.

#### 5. Study 2: Backfire effects for well-known vs. unknown brands

In Study 1, we did not observe any backfire effect of misinformation correction, even with the weaker "disputed" label. However,

the study focused on well-known global brands, which benefit from strong pre-existing associations that may shield them from misinformation. Lesser-known brands lack such anchoring, which could make consumers more vulnerable to rely on repetition-induced familiarity as a cue for a message's veracity. At the same time, the absence of prior background knowledge could also reduce the impact of repetition for claims about unfamiliar brands because consumers lack the memory structures (including unfavorable ones) that can make allegations feel more diagnostic and plausible. We investigate these competing possibilities by assessing the net effects for well-known and unknown brands. Study 2 employs a design almost identical to Study 1, except that half of the presented news stories featured unknown brands, while the other half featured well-known brands.

While in Study 1 corrections were found effective to counter the repetition effect, the overall effect of brand-related misinformation remains unclear, and thus, the question of whether correction can remedy the harm caused by being exposed to misinformation in the first place. Accordingly, Study 2 also includes a fourth 'default' condition (with no exposure to any news) to capture consumers' baseline brand evaluations. This allows us to assess the overall impact of misinformation on brand evaluations as well as to capture whether remnants of the misinformation effect remain even after correction with a strong cue ("false"). Study 2 is preregistered: <https://doi.org/10.17605/OSF.IO/5QXZA>.

### 5.1. Procedure and sample

Participants were first assigned to one of the four between-subject conditions (no-news vs. no-correction vs. "disputed" label vs. "false" label). In the no-news (default) condition, participants just evaluated 24 brands, 12 of which are well-known and 12 of which are hypothetical and unknown ("How do you feel towards the following brands" anchored by 1 = "Negative," and 7 = "Positive"). Then, they responded to demographic questions and terminated the study. Participants assigned to the three other conditions underwent a three-stage procedure identical to that of Study 1.

While the false and true news messages used in this study were identical to those in Study 1, half of the brands in those stories were converted to fictitious, unknown brand names. These news messages were designed with AI (OpenArt), which allowed us to keep the message content constant, while only manipulating the brand name. The rest of the procedure, as well as the last-stage labeling of misinformation (no-correction vs. "disputed" label vs. "false" label), remained identical to Study 1. The measures were also identical to those in Study 1. Only in the no-news (default) condition, participants did not answer the measures regarding belief in misinformation, readiness to share news, or news familiarity because they were not exposed to any misinformation in this condition. This study was administered to 1082 U.S.-based Prolific workers. After applying the pre-registered exclusion criteria, the sample comprised 987 participants ( $M = 40.26$  years,  $SD = 12.97$ ; 48% women;  $N = 169$  in the no-news, default condition, excluded from the main analyses).

### 5.2. Results

Two RM-ANOVAs estimated the repetition effect and the correction effect for the dependent variables (belief in misinformation, brand evaluations). The independent factors are the brand type (well-known vs. unknown, within-subjects), news type (true vs. false, within-subjects), exposure (repeated vs. novel, within-subjects), and the correction label (no-correction vs. "disputed" vs. "false" label, between-subjects).

**Modeling framework.** For belief in misinformation, RM-ANOVA does not provide a significant four-way interaction ( $F(2, 817) = 0.23, p = 0.798$ ). The main effects of the news type, exposure, and the correction label factors are significant (all  $ps < 0.001$ ), whereas the main effect of brand type is not significant ( $F(1, 817) = 0.09, p = 0.766$ ). Additionally, we observe two-way interactions between exposure and correction label ( $F(2, 817) = 4.34, p = 0.013$ ), news type and brand type ( $F(1, 817) = 14.82, p < 0.001$ ), and news type and correction label ( $F(2, 817) = 38.16, p < 0.001$ ). The interaction between news type and correction label is further moderated by the brand type ( $F(2, 817) = 5.74, p = 0.003$ ).

For brand evaluations, the four-way interaction is non-significant either ( $F(2, 817) = 0.46, p = 0.630$ ). The main effects of the news type, exposure, and brand type factors are significant (all  $ps < 0.001$ ), but the main effect of the correction label is non-significant ( $F(2, 817) = 0.40, p = 0.668$ ). Additionally, we observe an interaction between news type and brand type ( $F(1, 817) = 72.05, p < 0.001$ ), and three-way interactions among news type, brand type and correction label ( $F(2, 817) = 3.24, p = 0.040$ ) as well as among brand type, exposure and correction label ( $F(2, 817) = 4.63, p = 0.010$ ). We next analyze the contrasts of interest.

**Repetition effect for belief in misinformation.** Once again, the repetition of corporate misinformation increases the belief in it. This effect arises for both well-known brands ( $M_{\text{repeated}} = 2.51, M_{\text{novel}} = 2.21, \Delta_{\text{repetition}} = 0.30, p < 0.001, d = 0.45$ ) and unknown brands ( $M_{\text{repeated}} = 2.51, M_{\text{novel}} = 2.28, \Delta_{\text{repetition}} = 0.23, p < 0.001, d = 0.34$ ).

**Correction effect for belief in misinformation.** For well-known brands, the correction effect is significant when using the "false" label ( $M_{\text{no-correction}} = 2.51$  vs.  $M_{\text{false}} = 2.09, \Delta_{\text{correction}} = -0.42, p < 0.001, d = 0.64$ ) or the "disputed" label ( $M_{\text{disputed}} = 2.30, \Delta_{\text{correction}} = -0.21, p < 0.001, d = 0.32$ ). Both means of correction reduce consumers' belief in misinformation, but the "false" label does it more effectively.

For unknown brands, a similar pattern emerges: both labels reduce belief in misinformation, with the correction effect being more substantial for the "false" label. The magnitude of the correction effect of a "false" label remains similar to that of well-known brands ( $M_{\text{no-correction}} = 2.51$  vs.  $M_{\text{false}} = 2.07, \Delta_{\text{correction}} = -0.44, p < 0.001, d = 0.67$ ). The correction effect of the "disputed" label is weaker in the case of unknown brands but remains significant ( $M_{\text{disputed}} = 2.37, \Delta_{\text{correction}} = -0.14, p = 0.028, d = 0.21$ ).

**Net effect for belief in misinformation.** For well-known brands, the comparison between the repetition and correction effects produces a reversal effect for the strong ("false") label, such that the effect of correction exceeds that of repetition ( $\Delta_{\text{net effect}} = -0.12, p = 0.046, d = 0.19$ ), but only an offsetting pattern for the "disputed" label ( $\Delta_{\text{net effect}} = 0.09, p = 0.160, d = 0.13$ ). Although we do not identify a

**Table 5**  
Mean Comparison of Brand Evaluations to No-News Baseline.

Brand Type	Baseline (No-News)	vs. No-Correction	vs. Disputed Label	vs. False Label
Well-known	5.45	3.87***	3.88***	3.88***
Unknown	4.63	3.33***	3.30***	3.60***

Notes. Numbers represent the mean values of repeated misinformation.  $p$ -values reflect the contrasts to the baseline.

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

backfire effect, weak corrections can only compensate for the inflated beliefs in the piece of misinformation due to repetition-induced familiarity, while strong ones reverses these beliefs.

For unknown brands, the “false” correction label also achieves a reversal ( $\Delta_{\text{net effect}} = -0.21$ ,  $p < 0.001$ ,  $d = 0.33$ ). This is not the case for the “disputed” label’s correction effect which is, again, only offsetting the repetition effect ( $\Delta_{\text{net effect}} = 0.09$ ,  $p = 0.138$ ,  $d = 0.13$ ). Also for unknown brands, we do not obtain evidence for a backfire effect. Interestingly, the “false” label achieves a greater reversal for unknown brands than for the well-known brands, suggesting that unknown brands benefit more from stronger correction labels.

*Repetition effect for brand evaluations.* This effect is non-significant for both well-known ( $M_{\text{repeated}} = 3.87$ ,  $M_{\text{novel}} = 3.83$ ,  $\Delta_{\text{repetition}} = 0.04$ ,  $p = 0.549$ ,  $d = 0.04$ ) and unknown brands ( $M_{\text{repeated}} = 3.33$ ,  $M_{\text{novel}} = 3.28$ ,  $\Delta_{\text{repetition}} = 0.05$ ,  $p = 0.488$ ,  $d = 0.04$ ).

*Correction effect for brand evaluations.* For well-known brands, misinformation correction labels do not evoke a significant improvement in brand evaluations for both the “false” label ( $M_{\text{no-correction}} = 3.87$  vs.  $M_{\text{false}} = 3.88$ ,  $\Delta_{\text{correction}} = 0.01$ ,  $p = 0.952$ ,  $d = 0.01$ ) and the “disputed” label ( $M_{\text{disputed}} = 3.88$ ,  $\Delta_{\text{correction}} = 0.01$ ,  $p = 0.947$ ,  $d = 0.01$ ). However, for unknown brands, the “false” label improves evaluations of the brands victimized in the misinformation ( $M_{\text{no-correction}} = 3.33$  vs.  $M_{\text{false}} = 3.60$ ,  $\Delta_{\text{correction}} = 0.27$ ,  $p = 0.012$ ,  $d = 0.22$ ), while the “disputed” label does not ( $M_{\text{disputed}} = 3.30$ ,  $\Delta_{\text{correction}} = -0.03$ ,  $p = 0.761$ ,  $d = 0.03$ ).

*Net effect for brand evaluations.* For the well-known brands, the lack of significant repetition and correction effects results in non-significant net effects (“false” label:  $\Delta_{\text{net effect}} = 0.05$ ,  $p = 0.649$ ,  $d = 0.04$ ; “disputed” label:  $\Delta_{\text{net effect}} = 0.05$ ,  $p = 0.640$ ,  $d = 0.04$ ). By contrast, the benefit of correction is again stronger for the unknown brands. Accordingly, the “false” label results in a net improvement in brand evaluations ( $\Delta_{\text{net effect}} = 0.32$ ,  $p = 0.002$ ,  $d = 0.25$ ), which does not arise with the “disputed” label ( $\Delta_{\text{net effect}} = 0.02$ ,  $p = 0.884$ ,  $d = 0.01$ ).

*Overall label effectiveness.* We also sought to determine the overall impact of being exposed to the misinformation (with or without corrections) by assessing the brand evaluations in a condition where participants had not been exposed to any misinformation at all. We find that brand evaluations in this baseline condition are consistently higher than those in the conditions in which participants were exposed to the misinformation; this pattern even occurred for unknown brands (Table 5). This demonstrates that misinformation can harm the targeted brands. Our findings further imply that none of the correction efforts we tested improved brand evaluations to the level of the default brand evaluation condition. Evidently, even the strongest label (“false”) does not guarantee a full recovery from misinformation effects.

### 5.3. Discussion

Study 2’s results largely replicate those of Study 1 but reveal nuances pertaining to the role of brand familiarity. First, repetition increased belief in misinformation for both well-known and unknown brands, confirming the robustness of the illusory truth effect. The repetition effect was slightly stronger for well-known brands. This result may reflect greater cognitive fluency for familiar brands and will be discussed further in the general discussion.

Second, corrections significantly reduced belief in misinformation across brand types, with the “false” label consistently outperforming the “disputed” label. The “false” label not only offset the repetition effect but produced a reversal for both brand types. Yet for well-known brands, this net reversal effect was weaker: pre-existing brand associations may buffer against misinformation but also limit the impact of corrective efforts.

Third, concerning brand evaluations, repetition had no significant effect for either well-known or unknown brands. However, correction labels showed some impact. While they did not improve evaluations for well-known brands, the stronger “false” label led to a significant improvement for unknown brands. Yet, when compared to the no-news control condition, it becomes clear that exposure to misinformation, whether corrected or not, results in lower brand evaluations across both brand types. This indicates that even effective corrections do not fully restore brand perceptions to their baseline, underscoring the lasting reputational damage caused by misinformation, consistent with the continued influence effect (Lewandowsky et al., 2012).

Overall, we did not find evidence of a familiarity backfire effect across brand types, reinforcing the conclusion that issuing clear corrections is a safe strategy. Still, since even strong corrections do not fully restore brand evaluations to baseline levels, their benefits are limited.

## 6. Study 3: Correction when assessed in a later stage

Studies 1 and 2 examined backfire risk when correction labels are present at the time of assessment, such that consumers can use the corrective cue when reporting belief in misinformation and brand evaluations. Under these conditions, we observed no familiarity backfire effect. However, this does not establish whether corrective information is encoded and can be retrieved later when the label is

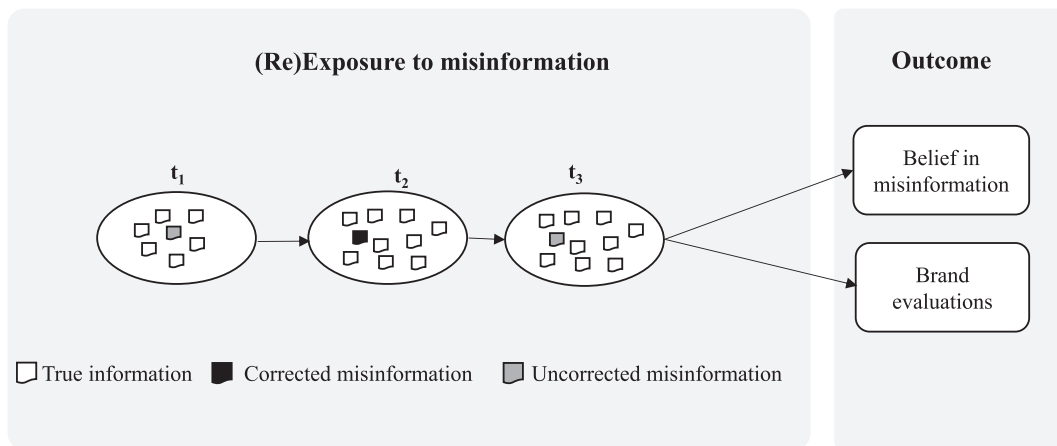


Fig. 2. Study 3 Procedure.

no longer available. A backfire effect could plausibly emerge in cue-absent contexts in which consumers must rely on memory, because familiarity with the repeated claim may persist while memory for the correction fades.

To address this possibility, Study 3 tests repetition and correction effects during a third, unlabeled exposure to brand-related misinformation following an earlier correction stage (see Fig. 2). This setting is managerially relevant because consumers receive no immediate guidance from a correction label at judgment, which should weaken correction effects, while an additional exposure further strengthens claim familiarity. Together, these features create a more conservative test of familiarity backfire, by increasing the relative weight of familiarity-based responding when correction cues are unavailable. Study 3 is preregistered: <https://doi.org/10.17605/OSF.IO/C7S5Q>.

### 6.1. Procedure

The procedure replicates that of Study 1, with two exceptions. First, given its relatively weaker effectiveness in earlier studies, we excluded the “disputed” label condition and only used the “false” correction label. Second, we added another stage to the experimental procedure (Fig. 2), such that the experiment comprises the initial exposure to 12 unlabeled stories (6 false, 6 true), a distractor task, a second exposure to the same 12 news stories with a “false” label in the correction label condition (vs. no-correction in the control condition), another distractor task, and finally a third, unlabeled, exposure to 24 unlabeled news stories (12 false, 12 true, half repeated and half novel), during which participants provided their belief in misinformation and brand evaluations. After the first (second) news exposures, participants reported their (their peers’) intentions to share the stories on social media. The materials consisted only of news stories about well-known brands, identical to those in Study 1.

The study thus features a  $2 \times 2 \times 2$  mixed design with the correction label manipulated between-subjects (“false” vs. no-correction), misinformation exposure (repeated vs. novel) and news type (true vs. false) manipulated within-subjects. The participants were 500 U.S.-based Prolific workers. After applying the preregistered exclusion criteria, the final sample consisted of 490 participants ( $M = 35.86$  years,  $SD = 12.53$ , 49% women).

### 6.2. Results

**Modeling framework.** For belief in misinformation, the RM-ANOVA shows a significant three-way interaction between the correction label, exposure, and news type ( $F(1, 488) = 42.91, p < 0.001$ ). All main effects and two-way interactions are significant (all  $ps < 0.001$ ). For brand evaluations, we neither observe significant main effects nor interaction effects, apart from an interaction between news type and correction label ( $F(1, 488) = 4.34, p = 0.038$ ).

**Repetition and correction effects for belief in misinformation.** Consistent with Studies 1 and 2, repeated misinformation exposure increases belief in misinformation ( $M_{\text{repeated}} = 2.33, M_{\text{novel}} = 2.14, \Delta_{\text{repetition}} = 0.18, p < 0.001, d = 0.36$ ). Also, the correction with a “false” label effectively reduces the participants’ belief in misinformation ( $M_{\text{no-correction}} = 2.33, M_{\text{false}} = 1.85, \Delta_{\text{correction}} = -0.48, p < 0.001, d = 0.92$ ). Bear in mind that the correction cue was not present when these assessments were made. Once again, we do not find evidence for a backfire effect provoked by the correction, but a reversal of the net effect. The magnitude of the correction effect exceeded the repetition effect ( $\Delta_{\text{net effect}} = -0.29, p < 0.001, d = 0.57$ ).

**Repetition and correction effects for brand evaluations.** While the repetition effect is again not significant in terms of brand evaluations ( $M_{\text{repeated}} = 3.54, M_{\text{novel}} = 3.59, \Delta_{\text{repetition}} = -0.05, p = 0.337, d = 0.05$ ), we observe a positive correction effect ( $M_{\text{no-correction}} = 3.54, M_{\text{false}} = 3.73, \Delta_{\text{correction}} = 0.19, p = 0.024, d = 0.20$ ). The “false” correction label improves consumers’ evaluations of the victimized brands appearing in repeated misinformation. However, contrasting the magnitudes of both effect types, the effect of correction does not reach significance in exceeding that of the repetition ( $\Delta_{\text{net effect}} = 0.14, p = 0.093, d = 0.15$ ). Yet, we again do not observe a backfire effect on brand evaluations.

### 6.3. Discussion

Study 3 tested for the backfire effect using a more conservative case, in which we reinforced the repetition effect (three exposures) and weakened the correction effect (correction label not present). When a correction label is present during the assessment (i.e., Studies 1 and 2), consumers can base their judgments on this external cue. In Study 3, we tested whether consumers actually integrate this cue into their memory structures and use it when being re-exposed to the misinformation without this guidance. By removing the correction cue from the assessment stage, we determined the strength of the repetition and correction effects in circumstances when consumers need to rely on their memory (i.e., having seen the correction label during previous encounters). Also in this case, misinformation correction appears effective in overcoming the repetition effect; the magnitude of the correction effect even significantly exceeded the repetition effect.

Correcting the misinformation also improved brand evaluations, while no harm in brand evaluations was registered due to the repetition of the misinformation, although the net effect did not reach significance. This finding for brand evaluations differs from the non-significant correction effects obtained in Study 1 and Study 2 (for well-known brands), where the judgments were formed while the correction cue was present. Hence, a stronger correction effect appears to arise for brand evaluations predominantly when consumers have to retrieve the cue from their memory (Study 3) or when unfamiliar brands are concerned (Study 2).

## 7. Study 4: Assessing the backfire effect for new audiences

### 7.1. Objective and study design

In the previous studies, we examined how consumers respond to misinformation correction after being initially exposed to false business-related claims without knowing they were inaccurate. In Study 4, we shift focus to a different, managerially relevant scenario. When a correction is issued, it often also reaches new audiences who encounter the misinformation for the first time, already accompanied by a corrective label. This situation raises an important question: does correcting misinformation at first exposure help prevent belief in false claims, or could it backfire? Prior research suggests that negated information tends to receive deeper cognitive processing, as individuals seek to understand the reason for the negation (Stanley et al., 2022). This increased attention can ironically strengthen belief in the underlying misinformation, especially when the correction lacks sufficient context (Autry & Duarte, 2021; Autry & Levine, 2012). This concern is particularly relevant in our research, as the misinformation targeting brands (sourced from real-world misinformation and corrected using brief fact-checking labels) often provides minimal explanation beyond a simple truth rating. Without elaboration or justification, these labels may fail to engage consumers in strategic processing, leaving them to rely on the familiarity of the claim itself. To address this, Study 4 investigates whether an initial exposure to misinformation with a fact-checking label leaves a memory trace that influences future encounters, particularly when the repeated exposure occurs without correction. This design allows us to assess whether early corrections benefit previously unexposed audiences or risk reinforcing the misinformation over time.

The  $3 \times 2 \times 2$  mixed design, the material, and the measures of this study are similar to those of Study 1, with one main change in the study design: misinformation is corrected during the first exposure (familiarization) but not during the repeated exposure. Hence, the correction condition (no-correction vs. “disputed” label vs. “false” label) is manipulated in the first stage of the experiment. A visualization of this design is provided in Web Appendix J. Study 4 is preregistered: <https://doi.org/10.17605/OSF.IO/TM9ES>.

This study was administered to 1573 U.S.-based Prolific participants. After applying the preregistered protocol, the final sample consisted of 1400 participants ( $M = 37.29$  years,  $SD = 13.80$ , 51% women).

### 7.2. Results

**Modeling framework.** For belief in misinformation, all main effects, two-way interactions, and the three-way interaction are significant (all  $ps < 0.001$ ). For brand evaluations, the model shows a significant main effect of the news type ( $F(1, 1397) = 35.86, p < 0.001$ ), an interaction of this factor with the correction label ( $F(2, 1397) = 3.35, p = 0.035$ ), and, most importantly, a significant three-way interaction with the exposure ( $F(2, 1397) = 4.27, p = 0.014$ ). All other effects are not significant (all  $ps > 0.198$ ).

**Repetition and correction effects for belief in misinformation.** Participants judge the repeated misinformation more accurate than the misinformation they had seen only once ( $M_{\text{repeated}} = 2.31, M_{\text{novel}} = 2.10, \Delta_{\text{repetition}} = 0.21, p < 0.001, d = 0.36$ ). Yet, the correction during the first exposure with a “false” label significantly reduces participant’s belief in the misinformation ( $M_{\text{no-correction}} = 2.31$  vs.  $M_{\text{false}} = 1.77, \Delta_{\text{correction}} = -0.54, p < 0.001, d = 0.95$ ). The more ambiguous “disputed” label also reduces belief in the misinformation ( $M_{\text{disputed}} = 2.05, \Delta_{\text{correction}} = -0.26, p < 0.001, d = 0.46$ ). Overall, the net effect of correction with the “false” label is stronger than that of repetition resulting in a clear drop in belief in the misinformation ( $\Delta_{\text{net effect}} = -0.33, p < 0.001, d = 0.58$ ), whereas the “disputed” label is only capable of offsetting the repetition effect ( $\Delta_{\text{net effect}} = -0.05, p = 0.144, d = 0.10$ ).

**Repetition and correction effects for brand evaluations.** The consequences of repetition for brand evaluations are not significant ( $M_{\text{novel}} = 3.68, M_{\text{repeated}} = 3.61, \Delta_{\text{repetition}} = -0.07, p = 0.093, d = 0.06$ ). Here, the correction with a “false” label helps maintain more positive brand evaluations compared to no-correction ( $M_{\text{no-correction}} = 3.61$  vs.  $M_{\text{false}} = 3.83, \Delta_{\text{correction}} = 0.22, p = 0.002, d = 0.21$ ), while the ambiguous “disputed” label appears ineffective in that it produces brand evaluations that are similar to those resulting from not labeling the misinformation at all ( $M_{\text{disputed}} = 3.65, \Delta_{\text{correction}} = 0.04, p = 0.527, d = 0.04$ ). As for the resulting net effect, we do not find evidence for a backfire effect. Both correction labels at least offset the (non-significant) effect of repetition, and the “false” label further improves brand evaluations (“false” label:  $\Delta_{\text{net effect}} = 0.16, p = 0.023, d = 0.15$ ; “disputed” label:  $\Delta_{\text{net effect}} = -0.02, p = 0.764$ ,

$d = 0.02$ ).

### 7.3. Discussion

Study 4 examined whether correcting misinformation at a first exposure, before any prior familiarity with the claim, can prevent belief in the misinformation or inadvertently backfire. This scenario reflects a common managerial reality. When firms issue corrections, they often reach new audiences encountering the misinformation for the first time.

The results show a clear repetition effect on consumers' belief in the misinformation, even when the initial exposure included a correction. Still, the "false" label significantly reduced this belief compared to no correction, and its effect was strong enough to reverse the repetition effect. The "disputed" label also reduced belief but only offset the repetition effect, without producing additional benefits. These findings suggest that even for new audiences, strong corrections are effective in reducing belief in misinformation and do not backfire.

For brand evaluations, the repetition effect did not reach significance, and only the "false" label improved evaluations compared to when the misinformation remains uncorrected. This improvement resulted in a net benefit beyond offsetting the repetition effect. The "disputed" label had no measurable impact. Overall, Study 4 confirms that issuing corrections during the first exposure is a safe strategy also for reaching new audiences. Here, strong fact-checking labels reduce belief in misinformation and help mitigate reputational harm.

### 7.4. Follow-up to Study 4: Delayed assessment of correction effects

As discussed in the theoretical background, the timing of belief assessment might play a critical role in determining whether consumers rely more on strategic processing or automatic familiarity when forming their judgments. Familiarity tends to persist over time, while memory for corrective details may fade, potentially increasing susceptibility to misinformation (Knowlton & Squire, 1995). In our previous studies, the delay between initial exposure and assessment was relatively short. To test whether a familiarity backfire effect might emerge over a longer interval, we conducted a preregistered follow-up study to Study 4 ( $N = 994$ ). Web Appendix J details this follow-up study.

This study replicates the multistage protocol of Study 4, with one key change: a three-day delay is introduced between the initial familiarization stage, during which misinformation is corrected, and the assessment stage. This design allows us to examine whether the effects of correction persist over time or diminish, potentially allowing repetition to dominate. The results show no evidence of a backfire effect, even after three days. Both correction labels reversed the (significant) effect of repetition on belief in misinformation but only offset the (non-significant) effect of repetition on brand evaluations.

## 8. Post-hoc analysis: Misinformation corrections' impact on true news

While the implications on true news stories are beyond the scope of this research, we still included a set of true news stories in our experiments for ecological validity. In real life, false and true stories tend to appear together on social media. We also conducted a post-hoc analysis of true news stories because these might be affected by correcting misinformation, too. For example, labeling can bias perceptions of surrounding unlabeled information, increasing their perceived accuracy (also termed "implied truth effect"; Pennycook et al., 2020). We assessed such spillover on brand-related true stories by comparing true news stories surrounded by corrected misinformation with true news stories surrounded by non-corrected misinformation. Again, we disentangle the effects of repeated exposure to misinformation and of correcting it. The results are presented in Table 6.

As Table 6 shows, we also observe repetition effects for true news, especially for belief. This is in line with prior research, suggesting that repetition inflates the truth judgment of various types of information (e.g., Udry & Barber, 2024). Additionally, we do not find an additional boost in such beliefs merely because of surrounding misinformation being corrected. That is, true stories surrounded by corrected misinformation were not perceived as more accurate than true stories surrounded by non-corrected misinformation. Similarly, none of the labels hint at "implied truth effects" for brand evaluations. Our findings thus diverge from those of Pennycook

**Table 6**  
Post-hoc Analysis of the Implications for Brand-related True Stories.

	Belief in True News			Brand Evaluations		
	Repetition Effect	Correction Effect ("Implied Truth Effect")		Repetition Effect	Correction Effect ("Implied Truth Effect")	
		Disputed	False		Disputed	False
<b>Study 1</b>	0.22***	-0.10 <sup>ns</sup>	-0.04 <sup>ns</sup>	-0.06 <sup>ns</sup>	0.04 <sup>ns</sup>	0.02 <sup>ns</sup>
<b>Study 2</b>						
Well-known brands	0.23***	0.03 <sup>ns</sup>	-0.08 <sup>ns</sup>	0.15*	0.01 <sup>ns</sup>	-0.09 <sup>ns</sup>
Unknown brands	0.18***	0.03 <sup>ns</sup>	0.01 <sup>ns</sup>	0.17**	-0.16 <sup>ns</sup>	-0.08 <sup>ns</sup>
<b>Study 3</b>	0.12***	-	-0.03 <sup>ns</sup>	0.04 <sup>ns</sup>	-	0.02 <sup>ns</sup>
<b>Study 4</b>	0.20***	-0.04 <sup>ns</sup>	-0.07 <sup>ns</sup>	0.03 <sup>ns</sup>	-0.06 <sup>ns</sup>	0.01 <sup>ns</sup>

Notes. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , ns: not significant ( $p > 0.05$ ).

et al.'s (2020) who observed an “implied truth effect” with “disputed” correction labels attached to political misinformation.

## 9. General discussion

This research addresses a critical question for both scholars and managers: can correcting misinformation about brands backfire by reinforcing the very claims it seeks to refute? While prior work has documented the illusory truth effect and raised concerns about familiarity-driven backfire effects, empirical evidence remains mixed and outside the marketing domain. Responding to calls for systematic investigation (Swire-Thompson et al., 2020; Walter & Tukachinsky, 2020), we disentangled, for the first time, the competing forces of repetition and corrections that shape the backfire effect. With five studies (and additional studies in the [Web Appendix](#)), we examined whether corrections can inadvertently increase belief in misinformation and explored their downstream consequences for brand evaluations.

We find robust evidence for the illusory truth effect for corporate misinformation with medium to large effect sizes. Repeated exposure consistently increased belief in misinformation. However, this effect was never strong enough to outweigh the benefits of correction. On the contrary, fact-checking labels, particularly those using unambiguous language (“False based on 3rd party fact-checkers”), effectively offset the repetition effect and, in several cases, reversed it. Importantly, we find no evidence of a familiarity backfire effect under any tested condition, including scenarios that the theory suggests should amplify its likelihood (e.g., unfamiliar brands, delayed assessment, or minimal corrective cues).

Across our studies, the unambiguous “false” label produced sizeable reductions in belief when misinformation was repeated (i.e., reversal rather than backfire). The corresponding net effects were moderate to large in most studies ( $0.33 \leq |ds| \leq 0.81$ ) with a smaller but reliable effect for well-known brands in Study 2 ( $d = 0.19$ ). In contrast, the ambiguous “disputed” label yielded net effects that were consistently close to zero across studies, indicating that it typically offsets repetition but does not reliably reverse it. Consistent with this pattern, across all studies, the correction effect of the “false” label was at least twice as large as that of the “disputed” label. For example, in Study 1, the correction effect on the original scale was  $-0.59$  (“false”) versus  $-0.18$  (“disputed”). We also did not observe stronger correction effects for unknown brands than for well-known brands. If anything, the correction effect was slightly stronger for well-known brands when the label was “disputed.” Together, these effects show that corrections, particularly unambiguous “false” labels, reduce belief without backfire, regardless of brand prominence.

For brand evaluations, the pattern differs markedly from belief in misinformation, although belief in the misinformation and brand evaluations were consistently negatively correlated ( $r_s = -0.083$  to  $-0.233$ ). Across all studies, repetition had no significant effect on brand evaluations, even for unknown brands, despite its robust effect on belief. Turning to correction, its impact on brand evaluations was generally weak and net correction effect sizes were often close to zero. Significant improvements emerged only in Study 2 for unknown brands and in Study 4 when the correction label was strong (“false”), with small net effects ( $d = 0.25$  and  $d = 0.15$ , respectively). In all other cases, corrections either had no effect or the effects were not strong enough to outweigh the (non-significant) repetition effect (Study 3). Importantly, even when correction improved evaluations, as in Study 2, it did not fully repair the damage: consumers’ brand evaluations remained substantially below the no-exposure baseline. While corrections can mitigate belief in misinformation, restoring brand equity is considerably more challenging.

While Studies 1 to 3 examined the typical scenario in which consumers first encounter misinformation and later see a correction, Study 4 and its follow-up focused on a different, managerially relevant situation; when corrections reach new audiences encountering the misinformation for the first time. This scenario raises concerns that corrections might inadvertently increase attention to the false claim, thereby strengthening its familiarity. Our results do not support this concern. Even when the initial exposure included a correction and the repeated exposure did not, strong labels (“false”) significantly reduced belief in misinformation and reversed the repetition effect, while weaker labels (“disputed”) provided smaller benefit by only offsetting it (Study 4) or reversing it with a small net effect (Follow-up to Study 4). The effect of corrections persisted even after a three-day delay, suggesting that corrections, especially strong ones, leave a durable memory trace that continues to protect against misinformation. For brand evaluations, we never observed a backfire effect, but the effects were modest: although strong corrections yielded improvements over no correction, these effects were more limited and only overturned the repetition effect when assessed immediately but not after three days.

In sum, fact-checking corrections do not trigger familiarity backfire effects, even under conditions designed to amplify such risks. While strong corrections reliably reverse repetition effects on belief, their impact on brand evaluations remains limited. Notably, our design incorporated several best practices recommended by Swire-Thompson et al. (2020). First, we minimized worldview-related confounds by focusing on commercial brands unlikely to evoke strong ideological positions and by randomizing which true and false stories participants saw, so not all participants encountered the same misinformation. Second, we used large samples across multiple experiments, reducing the risk of false positives. Third, although belief and evaluation were measured with single items per story, we averaged responses across multiple stories within each condition to avoid relying on a single observation. Together, these features strengthen confidence that the observed patterns reflect genuine effects rather than methodological artifacts.

### 9.1. Theoretical implications

Our findings align with prior research outside marketing that did not observe a familiarity backfire effect (e.g., Ecker et al., 2023; Swire et al., 2017) and refine theory by clarifying how three factors—type of corrective cue, brand familiarity, and timing of assessment—shape the competition between automatic familiarity (fluency-driven) and strategic, accuracy-seeking processing (Swire et al., 2017). Each factor can amplify or dampen these processes, thereby influencing whether corrections offset or reverse repetition effects, as discussed below.

Our data show that strong, unambiguous “false” labels reliably reverse the repetition effect on belief, whereas weaker “disputed” labels typically only offset it. These differences reflect meaning rather than perceptual attention: our pretests show both labels were seen as equally signaling untruthfulness, but “false” conveyed greater certainty and clarity (Web Appendix D/K), and an eye-tracking study indicates similar attention to both labels (Web Appendix E). This pattern suggests that higher cue diagnosticity creates cognitive friction at judgment, interrupting fluency-based familiarity and enabling strategic processing to dominate (Johnson & Seifert, 1994; Lewandowsky et al., 2012). By contrast, equivocal cues leave more room for fluency to guide judgments, consistent with reports of backfire under brief or weakly justified corrections (e.g., Skurnik et al., 2005; Thomas & Autry, 2024). Overall, our findings align with dual-process accounts (Swire et al., 2017) and extend meta-analytic evidence that more detailed debunking improves correction (Chan et al., 2017) to minimal fact-check labels; even within brief formats, stronger wording functions like “detail” by increasing diagnostic value and supporting strategic evaluation.

Concerning brand familiarity, the key result is that only for unknown brands did a strong correction label reverse the effect of repetition for both belief and brand evaluations at the same time. In all other cases (weaker labels and/or well-known brands), repetition and correction largely offset each other or reversed only the effect on belief but not on brand evaluations. This pattern is consistent with our reasoning that brand familiarity shifts the balance between repetition-based fluency and corrective cues. For belief, reversal for unknown brands reflects a weaker repetition effect combined with a strong correction effect; for well-known brands, repetition increased belief more strongly, consistent with the idea that richer brand schemas can make repeated allegations feel more diagnostic and plausible, thereby making fluency from repetition more likely to be taken as a truth cue (Hasher et al., 1977; Schwarz et al., 2007). For the evaluations, repetition was not significant for either brand type. For unknown brands, a strong correction nevertheless improved brand evaluations, plausibly because it provides a clear interpretive cue that the brand was unfairly targeted (Ecker et al., 2020; Lewandowsky et al., 2012). For well-known brands, evaluations remained comparatively resilient even when belief shifted, underscoring that reputational repair is harder than belief updating once misinformation circulates.

Concerning timing, even when corrective cues were absent (Study 3) or assessment was delayed (Study 4 follow-up), strong corrections continued to reduce belief and offset or reverse repetition effects, suggesting that corrective information was encoded and could be retrieved later without external cues. Taken together, these results support an integrated account in which corrections operate through cue-based evaluation when labels are present at judgment and through memory retrieval when labels are absent or delayed. Correction effects nevertheless attenuated under delayed, memory-based conditions, consistent with dual-process accounts in which recollection fades over time while repetition-driven familiarity remains relatively stable (Knowlton & Squire, 1995; Swire et al., 2017). Importantly, we still found no evidence of familiarity backfire, even under conditions theoretically conducive to it—multiple exposures, no corrective cue, and delay—consistent with reviews questioning its robustness (Swire-Thompson et al., 2020). For brand evaluations, timing mattered little: repetition remained non-significant, and corrections yielded only modest improvements, reinforcing that attitude repair is harder than belief updating (Lewandowsky et al., 2012).

Extending these insights, Study 4 (and follow-up) clarifies what happens when corrections reach new audiences at first exposure. Early corrections could either prevent misinformation from taking hold or inadvertently increase its salience; our results support the former. Strong labels substantially reduced belief, even after a delay, and did so about twice as much as weaker labels. This pattern is consistent with knowledge revision accounts in which corrections can leave a durable memory trace that counters familiarity as contextual cues fade (Ecker et al., 2017). Consistent with Ecker et al. (2020), we find no evidence that fact-check labels backfire by spreading misinformation to new audiences.

Notably, by systematically decomposing and quantifying the magnitude of both repetition and correction effects, our findings suggest that the importance of the illusory truth effect, though consistently observed, should not be overstated. Even the weaker “disputed” label reliably offset the repetition-induced boost in belief, indicating that minimal corrections can be sufficient to neutralize familiarity-driven distortions when appropriately framed.

## 9.2. Practical implications

This research underscores the managerial relevance of proactively addressing brand-related misinformation. The effects we observe are not confined to a single sector; they emerge across diverse industries, including technology, food, consumer goods, and automotive. Managers should be aware of both the risks and the safe practices identified in this research.

First, repetition increases belief. Across all studies, repeated exposure to misinformation systematically inflated belief, regardless of brand familiarity or the timing of the assessment. This repetition effect persisted even after a three-day delay, highlighting familiarity-based distortions. Importantly, the managerial risk of inflated belief extends beyond immediate shifts in broad brand ratings. Belief increases the perceived credibility of the claim and can guide downstream consumer responses, even when the reported attitudes are relatively stable (Ecker et al., 2022; Lewandowsky et al., 2012). When consumers treat an allegation as credible, they are more likely to incorporate it into their judgments and actions, such as, discounting the brand’s claims, updating perceived risk or trust, avoiding purchase or delaying consideration, and sharing or discussing the allegation with others (thereby amplifying reach). Consistent with this “continued influence” logic, Tybout et al. (1981) document that sales declined following brand misinformation even though respondents reported not believing the claim. In our data, belief in misinformation and brand evaluations were consistently related, but changes in belief do not fully account for consumers’ overall opinions about brands, especially for well-known brands anchored in strong prior attitudes.

Second, correction can be a safe and effective strategy. Despite widespread managerial concerns about potential backfire effects (cf. Pilot Study, Web Appendix A), our results show no evidence of such risks. On the contrary, corrections, especially clear and strong “false” labels, consistently reduced belief in misinformation. In Study 1, the correction effect of the “false” label was nearly four times

larger than the repetition effect, inflating belief, resulting in a 23.8% reduction in misinformation belief. Even the weaker “disputed” label offset repetition, but strong labels were more effective. Managers should feel confident issuing corrections rather than withholding them out of fear of familiarity backfire.

Third, restoring brand evaluations is more difficult. Even when corrections improved evaluations, as in Study 2, they did not fully repair the damage: final evaluations remained substantially below the no-exposure baseline. Because Study 2 is the only one that included a baseline condition, we cannot statistically test this pattern across studies; but a visual comparison of the corrected brand evaluations with the baseline attitudes consistently shows lower scores (Table 5). These results suggest that while corrections can mitigate belief in misinformation, restoring brand equity is considerably more challenging. Misinformation continues to influence downstream outcomes even after belief has been corrected (e.g., Ecker et al., 2010; Guillory & Geraci, 2016; Johnson & Seifert, 1994; Rich & Zaragoza, 2016). Although even unambiguous corrections cannot entirely restore brand evaluations, we urge managers to adopt those when correcting misinformation as they clearly outperform ambiguous ones.

Fourth, firms—not platforms—should take the lead in correcting misinformation. While social media platforms such as Facebook and X have previously implemented fact-checking labels, recent trends suggest they are scaling back these efforts (Silberling, 2025). Based on our findings, companies do not need to rely on platforms to intervene, but can instead initiate corrections themselves (Prike et al., 2023). This can be done by partnering with independent fact-checkers (e.g., PolitiFact, Snopes) who verify claims and issue labels that firms can use to publicly refute misinformation. Our results show that clear, unambiguous corrections are effective in reducing belief, even when misinformation is encountered for the first time. This supports a brand-led approach to misinformation management, where firms actively seek third-party validation and communicate it transparently to their audiences.

Fifth, clarity in labeling matters. Despite policy requirements for social media platforms to be more transparent in combating misinformation (Stano, 2020; EU Commission, 2022), platforms often soften fact-check labels (e.g., to avoid reactance, appear less paternalistic, protect engagement). Our findings suggest these concerns may be overstated: we found no evidence of label-induced reactance (Web Appendix D), and post-hoc analyses indicate users are more likely to share true than false news even without explicit corrections (Web Appendix L). Overall, audiences appear receptive to accurate content, so clear corrections—issued by platforms or firms—should help limit the spread and impact of false claims.

### 9.3. Limitations and future research

This research disentangled the competing effects of misinformation repetition and correction using fact-check labels. Prior work on other correction formats (e.g., “myth vs. fact”) sometimes reports longer-term backfire effects, especially in health contexts (e.g., Pluviano et al., 2019). Future research should test whether alternative formats yield different belief and evaluation outcomes for business-related misinformation. Another avenue concerns the broader information environment: political research finds that “disputed” labels can increase the perceived accuracy of nearby true stories (the implied truth effect; Pennycook et al., 2020). We did not replicate this effect (Table 6), leaving open whether implied truth effects are weaker in business settings, perhaps due to lower ideological involvement or more objective judgment criteria.

A further limitation is our focus on self-reported brand evaluations as the primary downstream outcome. As our results show, belief and evaluations are consistently related, yet experimental effects on evaluations are often small, especially for well-known brands with strong prior attitudes. This does not imply that misinformation lacks practical consequences. As illustrated by Tybout et al. (1981), marketplace outcomes can deteriorate even when consumers report disbelief, and the continued influence literature suggests that corrected misinformation can still shape subsequent inferences and decisions (Ecker et al., 2022; Lewandowsky et al., 2012). Future research should therefore test behavioral endpoints more directly (e.g., purchase/avoidance decisions, willingness to share or discuss the claim, complaint intentions) and examine whether familiarity-based backfire might emerge on behavior even when it is not detected on evaluation measures.

Although we tested diverse brands and misinformation topics (e.g., safety, health, discrimination), we did not examine how the emotional intensity of misinformation shapes outcomes. Prior work suggests that emotionally charged misinformation is more persistent and influential (Ecker et al., 2022; Thorson, 2016) and that emotions can bias truth judgments (Brashier & Marsh, 2020). Future research should test whether higher emotional intensity strengthens familiarity effects and makes belief distortions more persistent.

## 10. Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used Grammarly to improve the writing and OpenartAI to generate stimulus material for unknown brands in Study 2. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

### Financial disclosure

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft; DFG) [grant number: HO 5738/3-1] and the French National Research Agency (Agence Nationale de la Recherche; ANR) [grant number: ANR-18-FRAL-0012].

### CRediT authorship contribution statement

**Ipek N. Nibat:** Writing – original draft, Project administration, Visualization, Writing – review & editing, Investigation, Formal analysis, Validation, Methodology, Conceptualization. **Olivier Trendel:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Conceptualization, Formal analysis. **Robert Mai:** Writing – review & editing, Visualization, Supervision, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Tinka Krüger:** Writing – review & editing, Visualization, Resources, Conceptualization. **Wassili Lasarov:** Writing – review & editing, Conceptualization. **Stefan Hoffmann:** Writing – review & editing, Funding acquisition, Conceptualization, Resources.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijresmar.2026.03.007>.

### Data availability

Data will be made available on request.

### References

- Amazeen, M. A., Thorson, E., Muddiman, A., & Graves, L. (2018). Correcting political and consumer misperceptions. *Journalism & Mass Communication Quarterly*, 45(1), 28–48.
- Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, 48(1), 79–95.
- Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, 2(2), 81–94.
- Autry, K. S., & Duarte, S. E. (2021). Correcting the unknown: Negated corrections may increase belief in misinformation. *Applied Cognitive Psychology*, 35(4), 960–975.
- Autry, K. S., & Levine, W. H. (2012). Activation of negated and non-negated entities. *Journal of Pragmatics*, 44(11), 1474–1485.
- Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, 121(4), 446–458.
- Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, 47, 241–262.
- Berthon, P. R., & Pitt, L. F. (2018). Brands, truthiness and post-fact: Managing brands in a post-rational world. *Journal of Macromarketing*, 38(2), 218–227.
- Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology*, 71, 499–515.
- Business archives. Business Archives | Snopes.com. (n.d.). Retrieved 11/09/2025, <https://www.snopes.com/fact-check/>.
- Cameron, K. A., Roloff, M. E., Friesema, E. M., & Baker, D. W. (2013). Patient knowledge and recall of health information following exposure to “facts and myths” message format variations. *Patient Education and Counseling*, 92, 381–387.
- Carnahan, D., & Garrett, R. K. (2020). Processing style and responsiveness to corrective information. *International Journal of Public Opinion Research*, 32(3), 530–546.
- Cavazos, R. (2019). The economic cost of bad actors on the internet. *Technical Report*.
- Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66(3), 460–473.
- Chan, M. P. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546.
- Clayton, K., Blair, S., Busam, J. A., & Nyhan, B. (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.
- Cook, J., & Lewandowsky, S. (2012). *The debunking handbook*. St. Lucia, Australia: Queensland.
- Corneille, O., Mierop, A., & Unkelbach, C. (2020). Repetition increases both the perceived truth and fakeness of information: An ecological account. *Cognition*, 205, Article 104470.
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, 14(2), 238–257.
- Di Domenico, G., Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, 124, 329–341.
- Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, 38(8), 1087–1100.
- Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, 6(2), 185–192.
- Ecker, U. K. H., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2019). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology*, 111(1), 36–54.
- Ecker, U. K. H., Lewandowsky, S., & Chadwick, M. (2020). Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications*, 5(1), 41.
- Ecker, U. K. H., Lewandowsky, S., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psych.*, 1(1), 13–29.
- Ecker, U. K. H., Sharkey, C. X., & Swire-Thompson, B. (2023). Correcting vaccine misinformation: A failure to replicate familiarity or fear-driven backfire effects. *PLoS One*, 18(4), Article e0281140.
- European Commission (2022). Fighting COVID-19 disinformation. *Research report*. Retrieved 17/03/2026, <https://digital-strategy.ec.europa.eu/en/library/reports-march-and-april-actions-fighting-covid-19-disinformation>.
- Fishbein, M., & Middlestadt, S. (1995). Noncognitive effects on attitude formation and change: Fact or artifact? *Journal of Consumer Psychology*, 4(2), 181–202.
- Fong, J., Guo, T., & Rao, A. (2024). Debunking misinformation about consumer products: Effects on beliefs and purchase behavior. *Journal of Marketing Research*, 61(4), 659–681.

- Greifeneder R, Jaffe M, Newman E, & Schwarz N (2021). *The psychology of fake news: Accepting, sharing, and correcting misinformation* (Routledge).
- Guillory, J. J., & Geraci, L. (2016). The persistence of erroneous information in memory. *Applied Cognitive Psychology, 30*(2), 282–288.
- Hart, W. L., Albarracín, D., Eagly, A. H., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*, 555–588.
- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior, 16*(1), 107–112.
- Holan AD (2018, Feb 12). The Principles of the Truth-O-Meter: PolitiFact's methodology for independent fact-checking. PolitiFact. Retrieved 17/03/2026, <https://www.politifact.com/article/2018/feb/12/principles-truth-o-meter-politifacts-methodology-i/>.
- Johar, G. V. (2022). Untangling the web of misinformation and false beliefs. *Journal of Consumer Psychology, 32*(2), 374–383.
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1420–1436.
- Jussim, L., & Zanna, M. P. (2005). Accuracy in social perception: Criticisms, controversies, criteria, components, and cognitive processes. *Advances in Experimental Social Psychol., 37*, 1–93.
- Kemp, P. L., Alexander, T. R., & Wahlheim, C. N. (2022). Recalling fake news during real news corrections can impair or enhance memory updating: The role of recollection-based retrieval. *Cognitive Research: Principles and Implications, 7*(85), 1–21.
- Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 699–710.
- Kopalle, P. K., & Lehmann, D. R. (2015). The truth hurts: How customers may lose from honest advertising. *International Journal of Research in Marketing, 32*(3), 251–262.
- Lee, A. Y., & Labroo, A. A. (2004). The effect of conceptual and perceptual fluency on brand evaluation. *Journal of Marketing Research, 41*(2), 151–165.
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106–131.
- Mattavelli, S., Béna, J., Corneille, O., & Unkelbach, C. (2024). Repetition increases perceived truth of interpersonal statements from politically congruent and incongruent sources. *PsyArXiv*.
- Mende, M., Ubal, V. O., Cozac, M., Vallen, B., & Berry, C. (2023). Fighting Infodemics: Labels as antidotes to mis-and disinformation?! *Journal of Public Policy & Marketing, 43*(1), 31–52.
- Mosseri A (2017). Working to stop misinformation and false news. Retrieved 17/03/2026, <https://about.fb.com/news/2017/04/working-to-stop-misinformation-and-false-news/>.
- Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science, 66*(11), 4944–4957.
- Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021). A practical guide to doing behavioral research on fake news and misinformation. *Collabra:Psychology, 7*(1).
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General, 147*(12), 1865–1880.
- Peter, C., & Koch, T. (2016). When debunking scientific myths fails (and when it does not). The backfire effect in the context of journalistic coverage and immediate judgments as prevention strategy. *Science Communication, 38*, 3–25.
- Pluviano, S., Watt, C., & Della Sala, S. (2017). Misinformation lingers in memory: Failure of three pro-vaccination strategies. *PLoS One, 12*(7), Article e0181640.
- Pluviano, S., Watt, C., Ragazzini, G., & Della Sala, S. (2019). Parents' beliefs in misinformation about vaccines are strengthened by pro-vaccine campaigns. *Cognitive Processing, 1–7*.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research. *Journal of Applied Psychology, 88*(5), 879–903.
- Prike, T., & Ecker, U. K. H. (2023). Effective correction of misinformation. *Current Opinion in Psychology, 54*, Article 101712.
- Prike, T., Blackley, P., Swire-Thompson, B., & Ecker, U. K. H. (2023). Examining the replicability of backfire effects after standalone corrections. *Cognitive Research: Principles & Implications, 8*(1), 39.
- Rao, A. (2022). Deceptive claims using fake news advertising: The impact on consumers. *Journal of Marketing Research, 59*(3), 534–554.
- Rich, P. R., & Zaragoza, M. S. (2016). The continued influence of implied and explicitly stated misinformation in news reports. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(1), 62–74.
- Sanchez, J., Abril, C., & Haenlein, M. (2020). Competitive spillover elasticities of electronic word of mouth: An application to the soft drink industry. *Journal of the Academy of Marketing Science, 48*(2), 270–287.
- Schwarz, N., Jalbert, M., Noah, T., & Zhang, L. (2021). Metacognitive experiences as information: Processing fluency in consumer judgment and decision making. *Consumer Psychology Review, 4*(1), 4–25.
- Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy, 2*(1), 85–95.
- Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight. *Advances in Experimental Social Psychology, 39*, 127–161.
- Silberling A (2025, February 24). As Meta gets rid of fact-checkers, misinformation is going viral. *TechCrunch*. Retrieved 17/03/2026, <https://techcrunch.com/2025/02/24/as-meta-gets-rid-of-fact-checkers-misinformation-is-going-viral/>.
- Skurnik I, Yoon C, & Schwarz N (2007). Education about flu can reduce intentions to get a vaccination. Unpublished manuscript.
- Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31*(4), 713–724.
- Stanley, M. L., Whitehead, P. S., & Marsh, E. J. (2022). The cognitive processes underlying false beliefs. *Journal of Consumer Psychology, 32*(2), 359–369.
- Stano, S. (2020). *Routledge Handbook of Conspiracy Theories* (1st ed.). Routledge.
- Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1948–1961.
- Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition, 9*(3), 286–299.
- Thomas, J. G., & Autry, K. S. (2024). Unlicensed corrections violate the Gricean Maxims of Communication: Evidence for a cognitive mechanism underlying misinformation backfire effects. *Applied Cognitive Psychology, 38*(6), Article e70004.
- Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication, 33*(3), 460–480.
- Tybout, A. M., Calder, B. J., & Sternthal, B. (1981). Using information processing theory to design marketing strategies. *Journal of Marketing Research, 18*(1), 73–79.
- Udry, J., & Barber, S. J. (2024). The illusory truth effect: A review of how repetition increases belief in misinformation. *Current Opinion in Psychology, 56*, Article 101736.
- Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science, 28*(3), 247–253.
- Wahlheim, C. N., Alexander, T. R., & Peske, C. D. (2020). Reminders of everyday misinformation statements can enhance memory for and beliefs in corrections of those statements in the short term. *Psychological Science, 31*(10), 1325–1339.
- Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction. *Communication Research, 47*(2), 155–177.
- Wilson, E. A., & Park, D. C. (2008). A case for clarity in the writing of health statements. *Patient Education and Counseling, 72*(2), 330–335.