Fighting lies with facts or humor: Comparing the effectiveness of satirical and regular fact-checks in response to misinformation and disinformation

Boukes, M.; Hameleers, M.

DOI
10.1080/03637751.2022.2097284

Publication date
2023

Document Version
Final published version

Published in
Communication Monographs

License
CC BY-NC-ND

Link to publication

Citation for published version (APA):
Boukes, M., & Hameleers, M. (2023). Fighting lies with facts or humor: Comparing the effectiveness of satirical and regular fact-checks in response to misinformation and disinformation. *Communication Monographs*, *90*(1), 69-91. https://doi.org/10.1080/03637751.2022.2097284

COMMUNICATION MONOGRAPHS
2023, VOL. 90, NO. 1, 69–91
https://doi.org/10.1080/03637751.2022.2097284

OPEN ACCESS

Fighting lies with facts or humor: Comparing the effectiveness of satirical and regular fact-checks in response to misinformation and disinformation

Mark Boukes and Michael Hameleers

ASCoR, Universiteit van Amsterdam, Amsterdam, Netherlands

ABSTRACT

This study tested the effectiveness of fact-check format (regular *vs.* satirical) to refute different types of false information. Specifically, we conducted a pre-registered online survey experiment ($N = 849$) that compared the effects of regular fact-checkers and satirist refutations in response to mis- and disinformation about crime rates. The findings illustrated that both fact-checking formats – factual and satirical – were equally effective in lowering issue agreement and perceived credibility in response to false information. Instead of a backfire effect, moreover, the regular fact-check was particularly effective among people who agreed with the fact-check information; for satirical fact-checking, the effect was found across-the-board. Both formats were ineffective in decreasing affective polarization; it rather increased polarization under specific conditions (satire; agreeing with the fact-check).

ARTICLE HISTORY
Received 9 September 2021
Accepted 13 June 2022

KEYWORDS
Fact-checking; misinformation; disinformation; satire; fake news; credibility; polarization

Even though deliberative democracy revolves around the principle of a diverse and pluriform public sphere (Strömbäck, 2005), it is crucial that facts are still distinguishable from opinions: different political opinions should be founded upon the same factual reality (Arendt, 1967). In the setting of the current post-truth era (Bailey, 2018), the question on what action might be taken to correct misperceptions resulting from mis- and disinformation has become the subject of many empirical investigations (e.g., Wood & Porter, 2018). In this paper, we tested the impact of different interventions used to counter "fake news" – regular fact-checkers versus humorous refutations. The key question in this regard is whether the addition of humor to fact-checkers restricts its influence or helps to overcome the resistance of partisan-motivated reasoning.

Fact-checks have generally been found to be effective in correcting factual misperceptions (e.g., Amazeen et al., 2018; Nyhan et al., 2020; Walter & Tukachinsky, 2020; Wood & Porter, 2018). Yet, evidence on the effect of corrective information on partisan attitudes or political evaluations has been mixed at best (Nyhan et al., 2020; Walter & Tukachinsky, 2020). Especially among people who initially already supported the deceptive

CONTACT Mark Boukes markboukes@gmail.com
Supplemental data for this article can be accessed online at https://doi.org/10.1080/03637751.2022.2097284

claims of disinformation, the impact may be limited (e.g., Thorson, 2016). As satire's humorous format potentially overcomes people's tendency to critically scrutinize and counter-argue messages (Young, 2008), it may be regarded as an important journalistic tool to correct misinformation: accordingly, satire can be used to effectively hold politicians accountable (Richmond & Porpora, 2019) and even change the minds of their strongest supporters (Boukes & Hameleers, 2020). In this setting, we investigated whether satire can effectively correct mis- and disinformation, even among people whose beliefs align with the deceptive statements.

Although extant research compared the effectiveness of different modalities of fact-checking, such as visual rating scales or graphical information (Amazeen et al., 2018; Nyhan & Reifler, 2010), little is known about the impact of corrections provided in different genres of storytelling: we compared regular (i.e., purely factual) and satirical fact-checking articles in response to mis- and disinformation. Building on the work of Young et al. (2018), who compared a humorous video to a longer regular fact-checking article, we tested the effectiveness of text-only fact-checkers with and without humorous appeals. Although Young et al.'s findings indicate that both satirical and non-humorous videos can amplify the impact of fact-checking, we tested whether the impact of regular fact-checkers' presentation (written texts) can be amplified or weakened by incorporating humorous elements within the same modality (i.e., written content).

The current study investigated fact-checking in response to either false information framed with a deceptive political agenda (*disinformation*, often called "fake news") or false information without deceptive political agenda (*misinformation*). As disinformation refers to manipulated, doctored, or fabricated information created and disseminated with a political goal (Bennett & Livingston, 2018; Freelon & Wells, 2020), its intended impact is potentially more systematic and disruptive than misinformation. In addition, disinformation aims to increase polarization or cynicism by reaching specific audiences who are most susceptible to deceptive content (Erlich & Garner, 2021). Right-wing populists in Europe, for example, have mainly targeted deceptive anti-immigration disinformation to disenchanted native citizens, aiming to amplify societal cleavages (Bennett & Livingston, 2018). As disinformation is most effective among segments of the public that already support its claims (Schaewitz et al., 2020), it is relevant to consider to what extent the incorporation of humor in corrective messages could overcome resistance to corrections among vulnerable segments of the population.

To investigate the role of humor in corrective information, we conducted an online survey in which we compared the impact of regular fact-checking with fact-checking in a satirical format. We tested how effective both formats were to refute an article that falsely depicted increasing crime rate trends in general (misinformation without deceptive intent) versus a similar article that combined this information with an anti-immigration interpretation (reflecting the politics of right-wing disinformation). As central dependent variables, we considered effects of corrections on different outcomes: (1) perceived accuracy of misinformation, (2) issue agreement, and (3) de-polarized political attitudes. Motivated reasoning and confirmation biases – here understood as the guiding influence of people's prior beliefs on the persuasiveness of information (Festinger, 1957; Kunda, 1990) – are important to understanding the effectiveness of corrective information (Thorson, 2016), so we assessed whether prior beliefs that were (in)congruent with the corrected information moderate the fact-checking effect.

Misinformation and disinformation in political communication and journalism

Misinformation has been defined as information that is deemed incorrect based on the best available evidence and expert knowledge (e.g., Nyhan & Reifler, 2010; Vraga & Bode, 2020; Wardle, 2017). This study focused on two types of misinformation: factually incorrect information without clear political agenda (i.e., misinformation) versus deceptive information in which untrue information was attached to a political agenda (i.e., disinformation). With disinformation, political actors deliberately manipulate information to achieve political goals (e.g., Bennett & Livingston, 2018; Freelon & Wells, 2020; Wardle, 2017).

The dissemination of disinformation is mostly associated with (radical) right-wing issue positions; for example, cultivating anti-immigration support (e.g., Bennett & Livingston, 2018). Different from misinformation, disinformation is intentionally deceptive (Hancock & Bailenson, 2021). Such deceptive content aims to persuade recipients by misleading them – for example, by targeting their identities, emotions, and beliefs. Arguably, these intentionally deceptive messages are processed in a way that circumvents the detection of deception (Levine, 2014), as suspicion is not actively triggered when arguments resonate strongly with people's prior identities and beliefs (e.g., Thorson, 2016). As misinformation and disinformation have different (intended) consequences, we investigated the differential effectiveness of fact-checkers for both types of false information. In this setting, it was particularly relevant to explore whether false information that closely reflected a political agenda involving disinformation could be corrected with fact-checking information even among recipients that were likely to agree with the deceptive arguments.

Effects of regular fact-checkers and satirical fact-checkers on misinformation beliefs

Traditional fact-checking platforms, such as *PolitiFact.com* in the United States, *Africa Check* for African issues, *fullfact.org* in the United Kingdom, *Correctiv* in Germany, or *Poynter*'s international fact-checking network, check the veracity of political information by relying on empirical evidence, expert knowledge, and investigative journalism. Fact-checkers typically arrive at a verdict of the overall truthfulness of speeches, claims, and news articles: e.g., false, mostly false, mostly true, or true.

A growing body of research has investigated to what extent such fact-checks can successfully refute misinformation (e.g., Chan et al., 2017; Hameleers & Van der Meer, 2020; Nyhan et al., 2020; Thorson, 2016; Wood & Porter, 2018). Fact-checks may be effective in refuting misinformation because they combine simple and short messages with factual information, eventually reaching an unequivocal conclusion about a statement's truthfulness (Lewandowsky et al., 2012). Communicating short and factual counterarguments in response to false information should, ideally, result in the audience's acceptance of corrections (Chan et al., 2017). In that sense, fact-checks may "break through" the truth-default state of recipients by actively priming the idea of deception (Levine, 2014). In line with this, empirical research has found that fact-checks can be helpful in refuting untrue information, at least by correcting factual misperceptions (e.g., Nyhan et al., 2020).

Satire – defined as a ridicule or critique of human or individual vices, follies, abuses, or shortcomings by means of "a mixed bag" of humorous message types, such as irony, parody, or sarcasm (Holbert, 2013, p. 306) – has already been associated with the potential to correct misperceptions (Vraga et al., 2019; Young et al., 2018). Little is known, though, about the impact of satire on overcoming confirmation-biased processing of misinformation versus disinformation. Extant literature has suggested that the format of satire is very suitable to point out inconsistencies and false argumentation in political rhetoric (e.g., Boukes & Hameleers, 2020; Gaines, 2007; Richmond & Porpora, 2019; Waisanen, 2009; Warner, 2007). Research has also repeatedly shown that satire contributes to factual learning about political topics (e.g., Becker & Bode, 2018; Kim & Vishak, 2008; Young & Hoffman, 2012).

Satire should, however, be regarded as a genre that is less aligned with the traditional routines of journalism, such as striving for balance and facticity (Baym, 2005; Borden & Tew, 2007; Ödmark, 2021). Satirists are not bound to facticity – and are freer to actively and critically scrutinize the viewpoints of societal actors (Baym, 2005). Accordingly, satirical content may be suitable to hold politicians accountable (Boukes & Hameleers, 2020) and humorously highlights erroneous lines of argumentation or descriptions that are provided in misinformation with or without a strong ideological bias.

Thus, fact-checks and satirical refutations assign a different role to facticity: facts are central to the refutation strategy of fact-checkers, whereas satire's primary objective is to make a humorous appeal. Facts can be used instrumentally as a tool to point out the inaccuracies and fallacies of politicians' statements (Meddaugh, 2010), and thereby evoke laughter. Even though satire does not have to rely on actual facts, it mostly delivers solid argumentation in favor of or against certain (political) positions (Fox et al., 2007). Accordingly, both genres have the ability to correct misinformation: they offer a critique, ridicule or refutation of (political) issues and interpretations, and use arguments to raise suspicion about the presented misinformation.

Accordingly, Young et al. (2018) and Vraga et al. (2019) concluded that both humorous and non-humorous corrections can effectively refute misinformation. In their experiment, Young et al. (2018) found that videos (humorous as well as non-humorous) were more effective than textual fact-checkers. Within videos, however, using humor did not have an advantage compared to non-humorous videos. Vraga et al. (2019) additionally found that logic-based corrections were more credible among dismissive audience segments, whereas humor-based corrections were more effective among people convinced by false information.

As the next step, we assessed whether alternative types of fact-checking presented within the same modality (i.e., text) may also be effective to correct different types of false information (general misinformation vs. politicized disinformation). We tested this with the more generally consumed type of written satire, "parody news," which may be known from websites such as *The Onion* and *The Spoof* in the United States, the *Daily Mesh* and *News Thump* in the United Kingdom, *De Speld* in the Netherlands, and *Der Postillon* in Germany. The formats of these satirical platforms typically report on politics and public affairs with a satirical take beyond factual reality.

As fact-checks can lower the support for partisan positions that are strengthened by misinformation, corrections may de-polarize partisan cleavages (Hameleers & Van der Meer, 2020). Because corrective information should be most effective when there is room to correct misperceptions, oppositional political camps are potentially depolarized by promoting a common understanding of factual reality. Yet partisan polarization has been consolidated over a longer period of time, resulting from various gradual (dis)identification processes and selective exposure moments. In that sense, exposure to a single fact-check may not fully de-polarize existing partisan beliefs, but rather make these identities less salient and less powerful. Based on the aforementioned studies that looked at the effects of fact-checkers on misperceptions, corrective information should (a) lower the credibility of misinformation shown before the refutation, (b) lower the agreement with its "factual" claims, and (c) de-polarize the political issue attitudes of opposed-issue publics. This effect should be observed for both humorous and factual refutations (Young et al., 2018) compared to circumstances in which such as refutation of false information is not provided. Accordingly, we expected:

> H_{1abc}: Exposure to either a regular fact-checking article or to a satirical fact-checking article compared to the absence of a fact-checking article results in (H_{1a}) less issue-agreement with the claims made in misinformation, (H_{1b}) less perceived accuracy of the presented misinformation, and (H_{1c}) de-polarized political attitudes.

The differential effectiveness of regular versus satirical fact-checkers

The processing of fact-checking corrections may be subject to partisan interpretation (Nyhan & Reifler, 2010; Thorson, 2016). The concept of motivated reasoning resonates with the purpose of this study, because it explains why some people resist corrective information more than others. Motivated reasoning theory posits that people process information in a biased way to arrive at the most desirable conclusion; and subsequently tend to search for and accept arguments that confirm an already-supported position and avoid or reject arguments that challenge it (Kunda, 1990).

With (defensive) motivated reasoning (Festinger, 1957), people may either be motivated to arrive at accurate or consistent judgements. When people are defensively motivated, they process information in a way that defends their prior issue-beliefs to avoid the cognitive dissonance caused by incongruent information (Kunda, 1990). As corrective information may offer an attack on the existing political beliefs held by news users (Hameleers & Van der Meer, 2020), such information may cause cognitive dissonance. To avoid the corresponding sense of discomfort, people may be motivated to refute the attack of the correction and search their memory for cognitions that help them arrive at a desired conclusion that is consistent with their existing worldview (Darley & Gross, 1983).

Studies of fact-checks in the polarized political setting of the United States indeed found that exposure to fact-checks that counter the partisan views of Democrats or Republicans were less successful among people who supported the issue positions expressed in misinformation (e.g., Hameleers & Van der Meer, 2020; Nyhan & Reifler, 2010). When people are confronted with attitude-incongruent information in corrections, they may actually strengthen their preexisting agreement with misinformation instead of updating their misinformed beliefs in line with the fact-check. This is called reactance and a "backfire" effect: rather than accepting or rejecting the contradictory information, people may augment their existing beliefs in the face of an attack, as the challenging information limits people's perceived freedom (Miron & Brehm, 2006).

Even though Wood and Porter (2018) and Nyhan et al. (2020) did not find evidence for such a backfire effect, they found that political opinions beyond just factual perceptions are hard to correct. This is corroborated by the meta-analysis of Walter and Tukachinsky (2020): fact-checkers do not completely eliminate the effects of false information, and corrections are most successful when they confirm prior beliefs. Although abundant evidence is available that fact-checks can correct factually incorrect beliefs, evidence for the unconditional impact of fact-checks in correcting factual beliefs has been mixed at best. The effectiveness of fact-checks seems especially limited when trying to inform citizens across the ideological aisle.

Scholars have been more optimistic about the effectiveness of satirical formats to correct partisan misbeliefs (e.g., Richmond & Porpora, 2019). The humorous delivery of criticism in satire has been considered to present a well-argued counter-narrative (Hill, 2013). This is exactly what fact-checking also attempts to do: confronting the audience with an opposite – sometimes unexpected (i.e., counter-attitudinal) – view on a subject. Reactance is, however, less likely to occur following satirical messages than factual messages. First, satire is able to transport people into a storyline, which may lower the motivation to actively disagree with the message (Boukes et al., 2015; Nabi et al., 2007). Second, a relatively high amount of cognitive energy is required to fully comprehend satire, which implies that less cognitive capacity is available to counter-argue the message (e.g., Young, 2008), which should eventually overcome motivated reasoning processes during the consumption of satire.

It has indeed been found that satire may decrease support for political actors (Warner et al., 2018), and that especially initial supporters may change their minds (Becker, 2014). Boukes and Hameleers (2020), for example, found that humorous checks on the statements and promises of a populist party have a long-lasting and negative impact on the support for this party, with the strongest effects on supporters of the satirized party. In contrast to regular fact-checks, satire limits the motivation as well as the ability to counter-argue the correcting information from people's existing ideological pre-dispositions. Accordingly, satire may confront people with the truth without appearing too confrontational (Paletz, 1990) and, therefore, could be more effective than regular fact-checkers to also convince partisan citizens of "the truth."

Extending this argument, we expected that regular fact-checks – with their overly clear and straightforward message – would be relatively more successful in countering factual misperceptions compared to the more abstract and implicit fact-checks of satirical messages. But these satirist refutations, in turn, could be relatively effective in de-polarizing political beliefs compared to regular fact-checkers. Hence, we predicted that (a) issue-agreement and (b) perceived accuracy of the misinformation article were most strongly corrected by the factual arguments of regular fact-checks, whilst the less confrontational format of satire should result in a more effective correction of the harder-to-correct political beliefs measured as (c) depolarization in our study. We therefore proposed the following hypotheses on the relative effectiveness of regular fact-checking vis-à-vis satirical fact-checking:

> H_{2ab}: Exposure to a regular fact-check is more effective than exposure to a satirical fact-check to (H_{2a}) lower issue-agreement with the topic of the misinformation and (H_{2b}) lower perceived accuracy of the misinformation article.

> H_{2c}: Exposure to a satirical fact-check is more effective in de-polarizing political attitudes than exposure to a regular fact-check.

Confirmation-biased processing of fact-checking messages

False information may have the most detrimental effects on credibility, issue agreement, and political attitudes when people's prior (ideological) perceptions are reinforced (Hameleers & Van der Meer, 2020); and the correction of such false information may, subsequently, also be most difficult among these people. The mechanism underlying the rejection of fact-checking information may be understood as motivated reasoning (the confirmation-biased processing of information, see Knobloch-Westerwick et al., 2017; Winter et al., 2016). To fully understand the effectiveness of fact-checkers (satirical or not), one should consider attitudinal (in)congruence with the presented misinformation. As the experimental evidence offered by Nyhan and Reifler (2010) as well as Thorson (2016) indicated, fact-checks may not have the desired effects among partisans that agreed with the issue stances of misinformation (but see Wood & Porter, 2018): people with congruent partisan attitudes are generally less likely to update their political evaluations in line with a regular fact-checker's message (Nyhan et al., 2020). This led to the third hypothesis:

> H_{3abc}: The negative effect of exposure to a fact-checking article on (H_{3a}) agreement with the misinformation, (H_{3b}) perceived accuracy of the misinformation article, and (H_{3c}) polarization attitudes is stronger among participants who hold attitudinally-congruent perceptions of the fact-checking information compared to participants who hold attitudinally-incongruent perceptions of the fact-checking information.

Reactance was expected to be stronger for exposure to counter-attitudinal regular fact-checks than for counter-attitudinal satirical fact-checkers. After all, satire consumers are less likely to strive for or expect factual and balanced information (Feldman, 2007; Marchi, 2012). Rather, satire may be consumed for gratifications of entertainment (Diddi & LaRose, 2006).

Moreover, the humor in satire potentially weakens partisan-driven responses: opposed partisans are not necessarily perceived as the enemy (Jones & Baym, 2010) but as legitimate discussion partners (Paletz, 1990). Thus, satirical fact-checks may evoke the impression that the rejection is not an attack per se, but rather a gentle way to alert people about their misperceptions. Partisans may, accordingly, feel less offended when their predispositions are challenged by satire rather than regular fact-checking. Against this backdrop, we tested the following hypothesis on the role of prior issue agreement in response to the two formats of fact-checking:

> H_{4abc}: Participants with attitudinally-incongruent perceptions of the fact-checking information are more likely (H_{4a}) to lower their issue-agreement, (H_{4b}) to lower their perceived accuracy of the misinformation article, and (H_{4c}) to depolarize when exposed to a satirical fact-check compared to a regular fact-check.

While studying this topic, it was crucial to distinguish between erroneous information without an explicit political agenda (misinformation) and disinformation voicing a clear ideological or partisan agenda. Specifically, we expected that disinformation that articulates a clear political agenda (e.g., falsely connecting alleged increased crime rates to anti-immigration beliefs) may be more difficult to refute among people who strongly support the identity-based components of this deceptive message. Such information could strongly activate people's awareness of their partisan identity, which arguably produces a stronger motivation for confirmation-biased processing of fact-checking information that directly targets one's social identity (Kunda, 1990). Previous research demonstrated that a stronger resonance with prior attitudes made misinformation more persuasive and harder to refute (e.g., Hameleers & Van der Meer, 2020; Nyhan & Reifler, 2010; Thorson, 2016). Yet, political humor should be better able to correct such partisan-driven responses compared to a regular fact-based correction (e.g., Jones & Baym, 2010). The following moderation hypotheses were therefore tested:

> H_{5abc}: Participants with attitudinally-incongruent perceptions of the fact-checking information are (H_{5a}) less likely to lower their issue agreement, (H_{5b}) less likely to lower their perceived accuracy of the misinformation article, and (H_{5c}) less likely to depolarize after exposure to fact-checking information, and even less so for politicized disinformation compared to non-politicized misinformation.
>
> H_6: The moderating effect of attitudinal congruence is stronger for politicized disinformation than for non-politicized misinformation, and this moderating effect is stronger for a regular fact-checker than for a satirical fact-checker.

Method

Research design and procedures

A pre-registered online survey experiment was conducted in the *Qualtrics* environment. Details of this study (design, stimuli, hypotheses, and analyses) were pre-registered on OSF.[1] Following the recommendations of reviewers, we slightly adjusted the formulation of some hypotheses and included additional analyses for reasons of clarity or robustness.

By means of random assignment in equal group sizes, participants were first exposed to a news article that provided either the correct information (in the control condition), the misinformation, or the disinformation. Subsequently, participants received a correction (a regular fact-check or a satirical fact-check) or no refutation at all (a mock text with a non-related news item). Specifically, we employed a 3 (corrective information: control versus regular fact-check versus satirical fact-check) × 2 (misinformation: un-polarized vs. polarized) between-subjects factorial design. The additional control condition (a factually-correct news story on crime rate statistics followed by the control condition without fact-checker) was also part of the design but was not involved in the hypotheses and therefore does not return in the remainder of this manuscript. Figure 1 illustrates the design; all participants, thus, saw two messages.

Sample

Panel company *Dynata* was hired to collect data among a varied sample of U.S. participants early October 2020. We excluded participants living in Florida (to avoid interference with stimulus materials about the Miami-Dade County). A total of 850 participants completed the survey (1921 entered the survey, corresponding to a completion rate of 44.2%). The sample size was determined based on prior experimental research on fact-

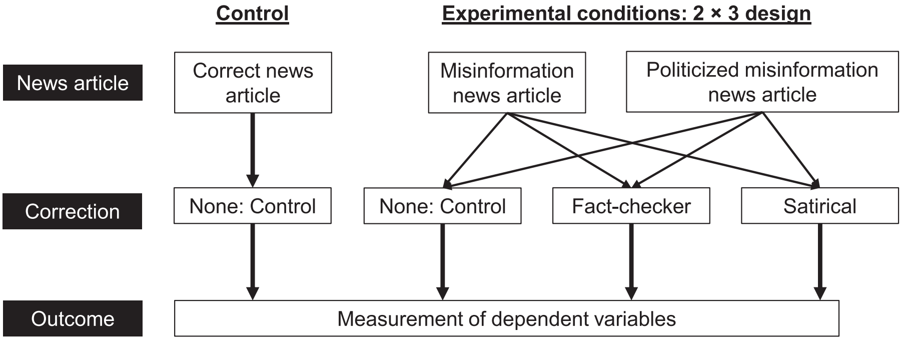
Control
Experimental conditions: 2 × 3 design
News article
Correct news article
Misinformation news article
Politicized misinformation news article
Correction
None: Control
None: Control
Fact-checker
Satirical
Outcome
Measurement of dependent variables

Figure 1. Visual depiction of experimental design and order of stimuli.

checking and misinformation and practical considerations (i.e., budget and timing). The relatively low completion rate was caused by the hard quota on age (18–91, *M* = 48.05, *SD* = 17.44), gender (51.1% female), education level (29.1% low, 43.2% moderate, 27.8% high) and partisan leaning (44.7% Democrat, 44.0% Republican, 11.3% Independent). No significant differences occurred across conditions in the sample composition on these key variables (i.e., randomization was successful).

Independent variables and stimuli

News Article: Misinformation vs. Disinformation. All stimuli were based on the template of an existing fact-checked United States newspaper article that emphasized the development of crime rate statistics in the United States.[2] The topic of increasing crime rates was chosen because many American citizens hold incorrect impressions about this issue (Gramlich, 2020). Based on this real-world template, there is consistent evidence that crime rates have decreased across all areas. We consciously decided to focus on crime rates in Miami-Dade – a region in Florida, a swing state, where it is not always clear which party is governing. The area has a considerable amount of undocumented Latino inhabitants, which offered a good opportunity for the manipulations of the two types of false information targeted in this paper: (1) a non-politicized misinformation version in which general crime rate estimates were falsely reported as an increase rather than a decrease and (2) a politicized disinformation version in which the same false statements were part of a broader right-wing political agenda where the increasing crime rates were attributed to undocumented Latino immigrants.

We used a neutral layout of an unknown news website and presented the article as an online news item. We avoided the use of source or party cues in layout and text to avoid unintended biasing effects (i.e., hostile media effect, see Boukes et al., 2014). The selected topic arguably resonated more with conservative than liberal partisan agendas, since we aimed to focus on cases of disinformation with high external validity. In line with this, most disinformation campaigns in the United States have been found to deliberately target polarized issue positions to sow discord or create anxiety on conservative issues, such as crime and immigration (Humprecht, 2019). All stimuli can be found in Appendix A.

Correction: Regular Fact-check vs. Satirist Fact-check vs. Control. For the regular fact-checking conditions, we used an existing fact-checker that refuted claims on increasing crime rates used by *PolitiFact* as a template. We made sure that the refutation was visible in the headline of the fact-check and used an actual visual rating scale known to be successful (Amazeen et al., 2018). The message was rated as "pants on fire," which means that all claims in the false information articles were completely untrue – for example, falsely stating that the politician was quoted. With minimal edits, we tailored the fact-checking conditions to the different false information conditions (e.g., mentioning "local citizens" or "Latino citizens"), and used the same verified crime-rate statistics to refute all claims of the false messages (also see Appendix A for fact-checking stimuli). For internal validity and comparability across conditions, we adapted the existing fact-checker of *PolitiFact* (both visually and textually) to realistically match the argumentation and empirical evidence of the specific false information conditions.

The regular fact-checking articles were transformed into the satirical fact-check: source cues were used from *The Onion* (a satirist platform in the United States) and funny quotes were inserted from an existing (but fictional) article in which the same political actor was accused of lying about crime rates – which was humorously connected to an alleged strategy of the actor "to gauge how much he'd be allowed to get away with."[3] Concretely, we integrated two funny quotes into the regular fact-checking stimuli to make the conditions as similar as possible in terms of provided content (i.e., internal validity): we removed all partisan cues and replaced them with neutral elements that matched the other stimuli (e.g., quotes were attributed to a fictional Florida lawmaker instead of to Donald Trump, as in the original material). We also added the original rating scale (i.e., "pants-on-fire") that was used in the fact-checking conditions. Participants in the control condition were not exposed to any fact-checking information. They were shown irrelevant information – hurricane news – that was of similar length to the fact-checking conditions. All texts were copy-edited multiple times by two independent native speakers.[4]

We used an existing source for the satirist fact-check (*the Onion*) and a realistic, but non-existing source for the regular fact-check (*PolitiCheck*). We used an existing source for the satirical conditions, because many people might otherwise not have recognized that they were exposed to a humorous message. In contrast, we decided to not use an existing fact-check, as conservatives tend to be more distrusting of these platforms, which could have biased our results. This was confirmed by our data: even before stimuli exposure, Democrats trusted fact-checkers ($M = 4.94$, $SD = 1.58$) more than Republicans ($M = 3.91$, $SD = 1.94$), $t(752) = 8.02$, $p < .001$, Cohen's $d = 0.58$. Yet, there were no differences in how familiar Republicans ($M = 6.82$, $SD = 2.99$) and Democrats ($M = 6.73$, $SD = 2.81$) were with the concept of "fact-checking" ($p = .673$).

Manipulation Checks. At the end of the questionnaire after the measurement of dependent variables, participants indicated how they perceived the two texts they had read (all 7-point disagree–agree scales). Statistical tests confirmed that manipulations of both the type of misinformation and the type of corrective information were successful. First, participants were more likely to recall (false) statements about increasing crime rates in the false information conditions ($M = 5.78$, $SD = 1.50$) compared to the control condition with the correct news article ($M = 3.00$, $SD = 2.15$), $t(848) = 17.59$, $p < .001$, Cohen's $d = 1.73$. The polarization manipulation was successful too: Whereas there were no differences in the correct recognition and memory of false statements on increasing crime rates across the two conditions with false information (mis- and disinformation; $p = .722$), participants in the polarized disinformation condition scored significantly higher on the correct recall of Latino immigrants being blamed for increasing crime rates ($M = 5.63$, $SD = 1.43$) compared to participants in the non-polarized misinformation conditions ($M = 3.68$, $SD = 2.08$), $t(728) = -14.76$, $p < .001$, Cohen's $d = 1.09$.

The manipulation of fact-checking presence was also successful: participants in the no-correction control condition scored significantly lower on the item "The second text was related to the first text" ($M = 2.95$, $SD = 2.95$) and "The second text stated that the original message was completely false" ($M = 2.96$, $SD = 1.99$) than participants in the corrective information conditions ($M = 5.43$, $SD = 1.66$ and $M = 5.40$, $SD = 1.69$, respectively: both $p < .001$). The more specific argument of the correction – arguing that crime rates have decreased rather than increased – was also associated more with the fact-checking conditions ($M = 5.29$, $SD = 1.77$) than with the control conditions ($M = 3.05$, $SD = 2.10$), $t(728) = 15.13$, $p < .001$, Cohen's $d = 1.19$.

Finally, participants correctly identified the format of the fact-check: they were more likely to associate the satirical fact-check ($M = 4.29$, $SD = 1.91$) than the normal fact-check ($M = 3.49$, $SD = 1.86$) with a satirist tone, $t(484) = -4.69$, $p < .001$, Cohen's $d =$ 0.43. Likewise, the regular fact-check was rated as significantly more serious ($\Delta M =$ 0.65, $p < .001$) and more earnest ($\Delta M = 0.54$, $p < .001$) than the satirist fact-check.

Measurements

The three dependent variables put forward in the hypotheses were measured immediately after exposure to the stimuli.[5] Items within the same scale were shown in a random order to avoid consistent question-order effects. If not stated otherwise, all items were measured on 7-point scales anchored by "completely disagree" and "completely agree" on the end-points. We calculated McDonald's omega (ω) as the measure of reliability, which produced very similar estimates as Cronbach's α, but better meets the statistical assumptions underlying its measurement model (Hayes & Coutts, 2020).

Issue-agreement. Participants indicated to what extent they disagreed or agreed with six statements on crime-rate developments that correspond with the content of news articles (McDonald's $\omega = .91$; $M = 3.01$, $SD = 1.67$; Skewness = 0.05, Kurtosis = −0.85): (a) the crime rates in the Miami area are increasing; (b) undocumented Latinos are responsible for increasing crime rates; (c) illegal aliens are responsible for causing violent crimes in our country; (d) governors should be more concerned about the increasing crime rates in our country; (e) the Miami area is a dangerous place because of increasing crime-related threats; (f) undocumented people with a Latino background are increasingly engaging in violent crimes.

Perceived Accuracy of Misinformation. Perceived accuracy of the news article on crime rates was measured with six items (McDonald's $\omega = .78$; $M = 2.70$, $SD = 1.36$; Skewness = 0.04, Kurtosis = 0.08): (a) the news item is Fake News (reverse-coded); (b) the news item tells the truth; (c) the news item is deceptive (reverse-coded); (d) the news item is accurate; (e) the news item describes the factual reality; (f) the news item is completely made up (reverse-coded); (g) the news item is based on erroneous information (reverse-coded).

Affective Polarization. We measured affective polarization as the difference between in-party liking and out-group hostility. Participants rated their feelings toward "individuals who support the Republican Party" and "individuals who support the Democratic Party" on a 10-point scale. Affective polarization was calculated as the difference between in- and out-party member ratings ($M = 3.90$, $SD = 3.96$; Skewness = −0.30, Kurtosis = −0.55); hence, scores could only be assigned to people who identified as Democrat or Republican, but not to those who identified as Independent ($n = 96$). The higher the score on the polarization variable, the stronger the difference between in-group and out-group liking (range −9 to +9). The survey also included two alternative measurements of polarization (i.e., affective candidate polarization and trait polarization). Their measurement and results are presented in Appendix B; overall, findings of these additional polarization scales are similar to those presented in the current manuscript.

Moderator: Attitudinal Congruence with Fact-checking Information. A battery of survey items was used – asked *prior* to stimulus exposure – to measure perceptions of increasing crime rates and the overall evaluation of immigrants (McDonald's $\omega = .80$; $M = 2.94$, $SD = 1.14$; Skewness = −0.04, Kurtosis = 0.16).[6] We did not describe the out-group as "Latinos" in these items, to avoid priming specific thoughts of this out-group before stimuli exposure. The scale consisted of six items: (a) the crime rates in our country are increasing (reverse-coded); (b) undocumented immigrants are dangerous (reverse-coded); (c) criminal acts are getting less common in the U.S.; (d) immigrants are involved in most violent crimes (reverse-coded); (e) undocumented immigrants cause more crimes than native U.S. citizens (reverse-coded); (f) our country is becoming more unsafe (reverse-coded). Items were (re)coded in such a way that they reflect the hypothesized variable of attitudinal congruence with the fact-checking information.

Data analysis

No missing data were present because participants were required to answer all questions to continue with the questionnaire. Following our pre-registration, we used a combination of analysis techniques to test our hypotheses and we used two-tailed significance tests for all analyses. First, a MANOVA showed that the three outcome variables were significantly dependent on the type of fact-check (control, regular, satirical) to which participants were exposed, $F(6, 1288) = 15.34$, $p < .001$, Wilk's $\Lambda = 0.87$.

One-way ANOVAs with Bonferroni *post-hoc* tests were then employed to examine the differences between the regular fact-checker and satirical fact-checking conditions versus the control conditions without fact-checking information (H_1). For the interpretation of differences, we reported the estimated marginal means for the different conditions in case of significant effects. The robustness of ANOVA results was verified – and confirmed – by additional two-way ANOVAs that controlled for the other independent factor in the experimental design (type of false information: mis- *vs.* disinformation). Next, independent samples t-tests were employed to test the differential effect of fact-checker format (H_2: regular *vs.* satirical).

The moderation hypotheses that predicted a conditional effect dependent upon the continuous variable "attitudinal congruence" ($H_{3,4,5,6}$) were tested with ordinary least squares (OLS) regression models, including the necessary interaction effects. To examine significant interaction effects, *Process* Model 1 (for two-way interactions) and *Process* Model 3 (for three-way interactions) were used to determine the regions of significance (i.e., Johnson–Neyman procedure) and to visualize the yielded interaction effects (Hayes, 2022).

Results

Hypothesis 1: Main effects of corrective information

One-way ANOVAs assessed the main effects of exposure to a fact-check message on lowering issue agreement (H_{1b}), perceived accuracy (H_{1b}) and de-polarization (H_{1c}).[7] In line with H_{1a}, we found support for a main effect of corrective information on issue agreement, $F(2,727) = 15.54$, $p < .001$, $\eta^2 = .04$. Specifically, compared to the absence of a correction in the control condition, exposure to a regular fact-checker ($M = 2.79$, $SE = .10$; $\Delta M = 0.78$, $SE = 0.15$, $p <$ 001) or satirist refutation ($M = 2.95$, $SE = 0.11$; $\Delta M = 0.62$, $SE = 0.15$, $p < .001$) caused significantly lower levels of issue agreement with the claims made in false information. This all supported H_{1a}.

We also found convincing support for H_{1b}. There was a significant negative main effect of exposure to fact-checking on perceived accuracy of the misinformation article, $F(2,727) = 52.33$, $p < .001$, $\eta^2 = .13$. This means that participants found the news article significantly *less* credible when it was fact-checked in a regular ($M = 2.20$, $SE = 0.08$; $\Delta M = 1.09$, $\Delta SE = 0.11$, $p < .001$) or satirical format ($M = 2.35$, $SE = 0.08$; $\Delta M = .94$, $\Delta SE = 0.12$, $p < .001$) compared to the absence of such corrective information ($M = 3.29$, $SE = 0.08$).

Regarding the impact of fact-checking on de-polarization (H_{1c}), we found no support for the third sub-hypothesis. Although the effect was significant, $F(2,646) = 3.02$, $p = .049$, $\eta^2 = .01$, it was in the opposite direction of what was expected. The pair-wise comparison of estimated marginal means with Bonferroni *post-hoc* correction revealed that exposure to the satirical fact-check resulted in *more* polarization compared to the control condition ($\Delta M = 0.92$, $\Delta SE = 0.38$, $p = .044$); no difference was found between the control and the regular fact-check ($p = .931$).

Hypothesis 2: Satirical fact-checking versus regular fact-checking

Next, we assessed whether the different formats of fact-checking affected the three outcome variables differently. We found no support for the hypotheses that regular fact-checks would be more effective than satirist refutations in lowering issue agreement (H_{2a}). Independent samples *t*-tests showed that fact-checkers ($M = 2.78$, $SD = 1.74$) were not significantly more or less effective than satirist refutations ($M = 2.95$, $SD = 1.68$) in lowering issue agreement with misinformation, $t(484) = -1.07$, $p = .284$.

Likewise, regular fact-checks ($M = 2.20$, $SD = 1.30$) were similarly effective to the satirical format ($M = 2.35$, $SD = 1.38$) in lowering the perceived credibility of misinformation, $t(484) = -1.21$, $p = .227$. Thus, H_{2b} was also not supported. Similarly, affective polarization was not affected differently by exposure to a regular fact-checker ($M = 3.83$, $SD = 3.96$) compared to a fact-checker of satirical format ($M = 4.37$, $SD = 3.89$), $t(484) = -1.44$, $p = .152$.

Thus, our findings did not offer any support for H_2: regular fact-checkers and satirist refutations were *equally* effective in (a) lowering issue agreement, (b) lowering perceived credibility, and (c) their effect on affective polarization was not significantly different.[8]

Hypothesis 3: The moderating role of prior attitudes for fact-checking's effect

Next, we tested the moderating effect of prior attitudes (H_3). Using an OLS-regression model, $F(3,726) = 205.81$, $R^2 = .46$, $p < .001$, we found a significant interaction effect between attitudinal congruence with the fact-check information and exposure to fact-checking information on the "issue agreement" dependent variable ($B = -0.17$, $SE = 0.07$, $p = .021$). In line with H_{3a}, this means that the negative relationship between fact-checking exposure and issue agreement with the false information was strongest among people who already agreed more with the fact-checking information (see Figure 2 for the plotted interaction effect). The graph shows that people who disagreed more with the fact-checking information (i.e., left-side of *x*-axis) did not significantly lower their issue agreement with the misinformation.

Looking at H_{3b}, there was no significant interaction effect between corrective information and attitudinal congruence on perceived accuracy of the misinformation ($B = -0.08$, $SE = 0.08$, $p = .303$), $F(3{,}726) = 52.62$, $R^2 = .18$, $p < .001$. This means that the perceived credibility of the false information article was reduced equally effective by the fact-checking information (compared to the control condition) among participants that initially already had congruent and incongruent attitudes on this topic. Thus, no support was found for H_{3b}.

Turning to affective polarization (H_{3c}), we found an insignificant interaction effect ($B = 0.48$, $SE = 0.25$, $p = .054$), $F(3{,}645) = 3.08$, $R^2 = .01$, $p = .027$. Zooming in on this result with the *Process*-macro (Hayes, 2022), though, we found a tendency in the data that

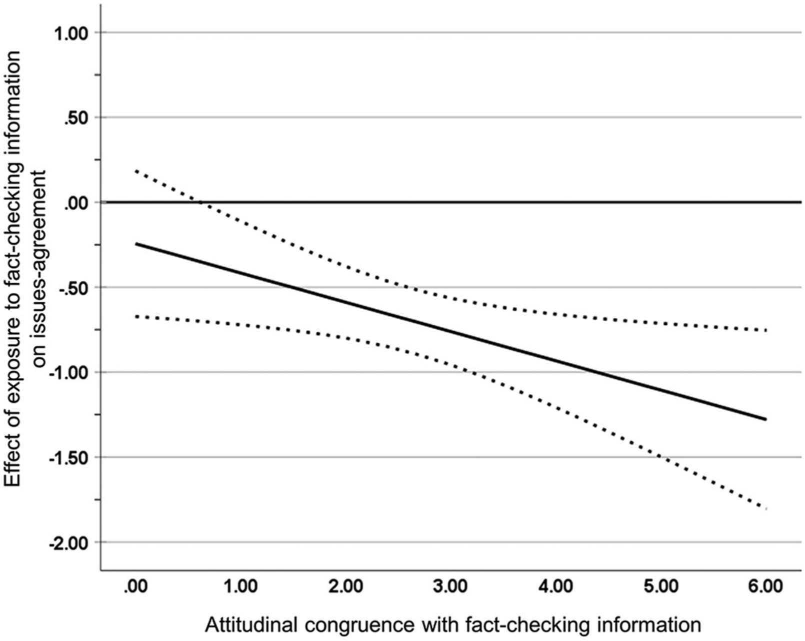

Figure 2. The visualized interaction effect on issue agreement (continuous line) and its 95% confidence interval (CI; dotted lines) of being exposed to fact-checking information (regular and satirical combined) relative to exposure to no fact-checking information for different levels of prior attitudinal congruence with the fact-check (*x*-axis: low to high congruence).

Note: When both sides of the CI are below the y-axis (effect on issue-agreement) at one point on the x-axis (attitudinal congruence), this indicates statistical significance of the effect on issue-agreement of exposure to fact-checking information at that particular level of attitudinal congruence. Datapoints are obtained with the Process-macro 4.0 (Hayes, 2022).

exposure to correcting information led to more polarization, but only among those with higher issue agreement (i.e., attitudinally congruent perceptions of the fact-checking information). This was in contrast with H_{3c}. Instead of lowering polarization, exposure to corrective information potentially strengthened polarization – especially among people with issue-congruent perceptions.

Hypothesis 4: Regular versus satirical fact-checking and attitudinal congruence

The moderating effect of attitudinal congruence was expected to be stronger for the regular fact-checker than for the satirist fact-checker. In support of H_{4a}, a significant interaction effect was found between attitudinal-congruence and exposure to the regular fact-check on issue agreement with the false information ($B = -0.25$, $SE = 0.09$, $p = .003$), whereas this interaction effect was insignificant for satire ($B = -0.09$, $SE = 0.09$, $p = .272$), $F(5,724) = 124.42$, $R^2 = .46$, $p = .027$.

Further analyses confirmed this pattern: the correcting effect of the regular fact-checker only occurred for people who already somewhat agreed with the fact-check. The effect of the satirical fact-checker, however, occurred for all participants. This confirmed H_{4a}. Yet, it should be noted that the difference between both interaction effects was not significant in itself ($p = .064$). So we only found partial evidence that the moderating effect of attitudinal-(in)congruence was stronger for the regular fact-checker than for the satirical fact-check; less motivated reasoning might occur after exposure to a satirical fact-check, causing an effect across the board.

Looking at perceived accuracy (H_{4b}), we found no difference in the effectiveness of the two different fact-checking types for people who held more or less congruent attitudes with the correcting information. The interaction effects between congruence and either the regular fact-check ($B = -0.111$, $SE = 0.09$, $p = .208$) or the satirical fact-check ($B = -0.04$, $SE = 0.09$, $p = .639$) were both insignificant, $F(5,724) = 31.82$, $R^2 = .18$, $p < .001$. Thus, H_{4b} could not be supported.

For the interaction effects between prior attitudes and different formats of fact-checking on affective polarization (H_{4c}), we found that there was a significant interaction effect for the regular fact-checker ($B = 0.71$, $SE = 0.29$, $p = .014$), but not for the satirical fact-checker ($B = 0.29$, $SE = 0.29$, $p = .304$), $F(5,643) = 2.78$, $R^2 = .02$, $p = .017$. Hence, exposure to a regular fact-check increased polarization, but only among participants that agreed with the fact-checker. This moderation was not the case for satire; exposure to the satirical fact-checker increased polarization for all people. All in all, H_{4c} was not supported: although we did confirm the overall patterns that satire worked across the board, the effects of fact-checks were contingent upon the resonance with prior attitudes (H_4). The effects consistently pointed in the *opposite* direction of our hypothesis (i.e., fact-checking increased rather than decreased polarization).

Hypothesis 5: Fact-checking for misinformation and disinformation

We expected attitudinal congruence to play the strongest role in polarized disinformation (H_5). Contrary to our expectations, we found no difference in the moderating role of attitudinal congruence on issue agreement for unpolarized misinformation versus polarized disinformation. The three-way interaction effect between (a) exposure to a fact-check or not, (b) exposure to unpolarized or unpolarized misinformation, and (c) attitudinal congruence with the fact-check was insignificant, $b = 0.07$, $SE = 0.15$, $p = .651$, $F(7{,}722) = 92.24$, $R^2 = .47$, $p < .001$. H_{5a} was therefore not supported. The same was found for H_{5b}: there was no significant three-way interaction effect between the three factors on the perceived accuracy of the false information ($p = .223$).

We did, however, find a significant three-way interaction effect for the interaction between attitudinal congruence, corrective information, and type of misinformation on affective polarization ($B = 1.29$, $SE = 0.50$, $p = .010$), $F(7{,}641) = 2.64$, $R^2 = .03$, $p = .011$. This interaction effect is visualized in Figure 3, which shows that exposure to a fact-check generally led to more polarization. The only exception was the polarized disinformation condition among people who held attitudinally incongruent perceptions of the fact-check. They may have realized that their previous opinions were wrong or too extreme, and may have corrected their polarization perceptions after fact-check exposure.

Affective polarization was triggered most strongly, in contrast, by exposure to the fact-check among people who already agreed with the fact-check (high congruence) under the condition of polarized disinformation. This all contradicted H_{5c}, which predicted that depolarization would be least likely among people with low attitude congruence after fact-checking polarized disinformation. Our findings, thus, did not offer support for H_5.

Hypothesis 6: Regular and satirical fact-checking for mis- and disinformation

Finally, we turned to H_6. We expected that the moderating effect of attitudinal congruence would be stronger for politicized disinformation than for non-politicized misinformation, and that this moderating effect would be stronger for a regular fact-check than a satirist fact-check. Our data did not support this hypothesis; the three-way interaction effects between issue-congruence, type of false information, and fact-check format

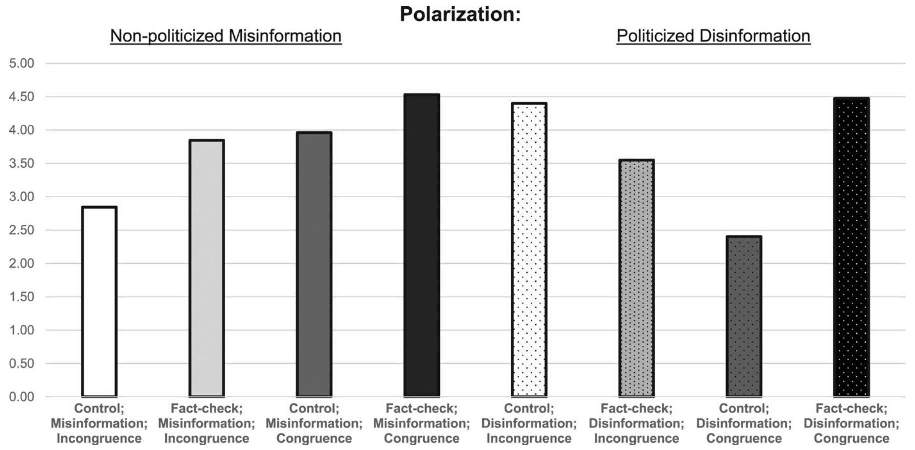

Figure 3. Mean polarization scores for the presence and absence of a fact-check message at different levels of attitudinal congruence with the fact-check (+1 *SD* above and −1 *SD* below the mean of attitudinal congruence) – specified for unpolarized misinformation and polarized disinformation conditions.

were not significant for issue agreement ($p = .958$), perceived accuracy ($p = .297$) or affective polarization ($p = .799$).

Summary of findings

An overview of all findings is presented in Table 1.

Discussion

Regular and satirical fact-checks have successfully lowered issue agreement with false information and reduced its perceived accuracy. However, both types of fact-checking have not been effective in de-polarizing political attitudes. If anything, satirical fact-checks may increase rather than decrease the level of polarization. A similar pattern has also been detected in research on selective exposure (Stroud & Muddiman, 2013) where satire may turn news consumers away from oppositional views, and thus reduce the tolerance toward incongruent political views. Our experiment shows that a similar response can be caused by satirical corrections: instead of the intended impact of overcoming polarization, using satire in corrective information strengthens existing negative political evaluations of the opposite party. Therefore, our findings point to a worrisome side-effect of corrective information. It may reinforce the partisan beliefs of people

Table 1. Overview of findings.

		Dependent variable	
Independent variable	**Issue-agreement**	**Perceived accuracy**	**Affective polarization**
Control *vs.* Fact-check:			
H_1: Fact-check exposure	Less agreement with false information	Lower perceived accuracy of false information	More polarization (but only significant for satirical fact-check)
H_3: Fact-check × Attitudinal congruence	Fact-checking more impact full for people agreeing with the fact-check	n.s.	Fact-checking causes more polarization among people who agree with the fact-check (but $p = .054$)
H_5: Fact-check × Mis- vs. Disinfo × Attitudinal congruence	n.s.	n.s.	Fact-check exposure generally leads to more polarization, but not for disinformation among people with an incongruent attitude regarding the fact-check.
Regular fact-check *vs.* Satirical fact-check:			
H_2: Regular vs. Satirical fact-check	n.s.	n.s.	n.s.
H_4: Regular vs. satirical × Attitudinal congruence	Negative effect of satirical fact-check occurs across the board; Regular fact-checking only influenced people with congruent attitude	n.s.	The satirical fact-checker increased polarization for all people; Regular fact-checker increased polarization only among participants with a congruent attitude.
H_6: Regular vs. Satirical × Mis- vs. Disinfo × Attitudinal congruence	n.s.	n.s.	n.s.

Note: Description of findings is only included for significant relationships ($p < .050$). Non-significant difference are denoted with n.s., which are therefore not interpreted in terms of effect patterns.

already aligning with the correction but may have no effect on citizens who really need fact-checking information to correct their existing misperceptions.

Our main findings mostly correspond with existing fact-checking literature (Hameleers & Van der Meer, 2020; Nyhan et al., 2020; Wood & Porter, 2018). An important contribution to the literature is the finding that different formats of fact-checking information do not necessarily strengthen or weaken the overall impact of corrective information: a satirical format of fact-checking can be used alongside regular factual corrections to lower issue-agreement with the misinformation and challenge the credibility of fake news articles. Thus, our findings show that media practitioners can combat misinformation via different routes.

We have found that corrective information is most effective for participants who already hold an attitude congruent with the fact-checking message – in our study, people who were aware that crime was not increasing and had fewer negative perceptions of immigrants. Unfortunately, the fact-checks have not significantly corrected the issue-agreement of people who initially aligned most with the false information. Although we did not find a backfire effect when looking at issue agreement and credibility of misinformation in response to false information (see also Nyhan & Reifler, 2010), we did find that corrections are more likely to be accepted when fact-checks confirm prior beliefs. Yet we should be careful about drawing too strong causal conclusions about the conditional effects because the moderator in our study is an observed variable (i.e., attitudinal congruence) that was not randomly assigned to participants.

Additionally, the format of fact-checking contributes to its effect. More specifically, where regular fact-checkers only succeed in convincing people who already agreed with the fact-checker's stance, satirical fact-checking has a more universal effect. This shows that satirical refutations may be less susceptible to resistance and confirmation biases – a finding that is also reflected in previous research on the power of satire (Boukes & Hameleers, 2020; Young et al., 2018). By transporting audiences into its narrative (Boukes et al., 2015; Nabi et al., 2007) and requiring a high cognitive load (Young, 2008), satirical fact-checking information produces less resistance in the processing of it message. The satirical fact-check, however, also increased affective polarization across the board, whereas following regular fact-checking, this only happened among those who saw their existing views confirmed by the fact-check.

We should note that the effects of the different corrections are not the same for all three outcome variables. Although it could be argued that the aim of corrective information is, firstly, to lower the perceived accuracy and credibility of false information, fact-checking may also have the secondary purpose of reducing polarization along partisan lines. Our findings indicate that depolarization is hard to achieve and fact-checking even seems to cause the opposite. We find that when prior attitudes align more with the fact-checking information, stronger polarizing effects do occur. One potential explanation is that political worldviews and social identifications are reinforced among people agreeing with the fact-checker, where the fact-checker offers evidence about the negative traits of the partisan out-group ("they are spreading fake news") while reassuring a positive in-group identity ("they are wrong, we are right"). Among people opposing the fact-checking information (i.e., those whose prior attitudes align with misinformation), more doubt about their own identifications and political worldviews may arise due to the fact-checker's attack on their beliefs. Future research is needed to disentangle the exact processing mechanisms at play. This research can also further explore the structural – and potentially mediating – relationships between these different outcome variables.

Our experiment is not without limitations. First, we have focused on a single topic – crime. While we have assessed the differential effects of unpolarized misinformation versus polarized disinformation, crime may already be seen as a politicized issue in itself – which potentially explains the lack of effects caused by this manipulated factor. People could have formed stable attitudes on this issue that are hard to change, especially with a single message. Future research may test the same processes using a diversity of issues that are perceived as less versus more important and less versus more politicized. As we have looked at an issue that is generally more aligned with conservative than liberal issue positions, the question also remains how transferable our findings are to issues with an implicit liberal agenda, such as misinformation on climate change or LGBTQIA+ rights.

More research is also needed that investigates the impact of disinformation and corrections outside of the United States, especially in countries with a larger set of represented political parties where audiences are polarized on multiple dimensions (Freire, 2015). Even in such multi-party contexts, however, polarized divides can be identified across issues, such as immigration, climate change, or vaccination, while not necessarily being related to a specific ideology or political party. Similar confirmation biases may thus play a role there, and the challenges of [delivering, examining, producing] impactful fact-checking may be similar. Therefore, we suggest others replicate our findings in different national contexts to verify the robustness of our conclusions.

In addition, the current study has investigated the impact of fact-checking in the short term, while neglecting the duration of effects and the embeddedness of corrective information in the wider (digital) media environment. One can, for example, expect that partisan beliefs and polarization are long-term processes not easily affected by randomly exposing people to a single fact-check. By isolating responses to a single exposure moment, we could tease out the processing of corrected mis- and disinformation, but we cannot arrive at a realistic assessment of how the over-time integration of corrections in partisans' newsfeed may influence polarized divides.

Finally, we conclude that regular and satirical fact-checking are equally effective in correcting the misperceptions resulting from false information, but at the same time may increase affective polarization. More specifically, the beneficial (correcting misperceptions) and detrimental (polarization) effects of regular fact-checking are most likely among the people with attitudes congruent to the fact-check. These effects – also on increased polarization – occurred unconditionally (i.e., across the board) for satirical fact-checking.

Notes

1. See: https://osf.io/5xrdw/?view_only=f45fed93ce3d476d837a895b28072037
2. See: https://www.politifact.com/factchecks/2019/jul/25/steve-mccraw/how-much-has-crime-gone-down-texas-mexico-border/
3. See: https://politics.theonion.com/trump-planning-to-throw-lie-about-immigrant-crime-rate-1819579272

4. We thank Eugenia Quintanilla (University of Michigan; Ann Arbor) and Paris Bethel for their support on this.
5. The dependent variables did not correlate strongly with each other: Issue-agreement and perceived accuracy ($r = .37$, $p < .001$); issue-agreement and affective polarization ($r = -.13$, $p < .001$); perceived accuracy and affective polarization ($r = -.07$, $p = .068$). The three variables, accordingly, most likely reflect different theoretical constructs.
6. The moderator scale was not clearly operationalized in the pre-registration. Although the survey scale included two additional items ("Immigrants contribute to the cultural richness of our country" and "Immigration provides high-skilled workers on our labor market"), these were not included in the scale because both do not resonate explicitly with the presented (and fact-checked) news item.
7. One-way ANOVAs were conducted following our pre-registration. Per the reviewer's suggestion, we verified the robustness of our results with a two-way ANOVA that controls for the type of false information (mis- vs. disinformation). All reported effects remained significant and in the same direction.
8. These findings are confirmed in two-way ANOVAs that also control for the effect of mis- versus disinformation.

Acknowledgements

This study received funding from the Replication Grant from the Amsterdam School of Communication Research (ASCOR) fund for Open Science in 2020. Moreover, this work was supported by the Dutch Research Council (NWO) with a Veni grant under project number: 016.Veni.195.123 awarded to Dr Mark Boukes.

Disclosure statement

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

Funding

This work was supported by Replication Grant from the Amsterdam School of Communication Research (ASCOR) fund for Open Science in 2020 and Dutch Research Council (NWO) with a Veni grant under project number: 016.Veni.195.123 awarded to Dr Mark Boukes.

ORCID

Mark Boukes http://orcid.org/0000-0002-3377-6281
Michael Hameleers http://orcid.org/0000-0002-8038-5005

References

Amazeen, M. A., Thorson, E., Muddiman, L., & Graves, L. (2018). Correcting political and consumer misperceptions. *Journalism & Mass Communication Quarterly*, *95*(1), 28–48. https://doi.org/10.1177/1077699016678186

Arendt, H. (1967). Truth and politics. *The New Yorker*. https://www.newyorker.com/magazine/1967/02/25/truth-and-politics.

Bailey, R. (2018). When journalism and satire merge: The implications for impartiality, engagement and 'post-truth' politics. *European Journal of Communication*, *33*(2), 200–213. https://doi.org/10.1177/0267323118760322

Baym, G. (2005). The daily show: Discursive integration and the reinvention of political journalism. *Political Communication*, *22*(3), 259–276. https://doi.org/10.1080/10584600591006492

Becker, A. B. (2014). Playing with politics: Online political parody, affinity for political humor, anxiety reduction, and implications for political efficacy. *Mass Communication and Society*, *17*(3), 424–445. https://doi.org/10.1080/15205436.2014.891134

Becker, A. B., & Bode, L. (2018). Satire as a source for learning? *Information, Communication & Society*, *21*(4), 612–625. https://doi.org/10.1080/1369118X.2017.1301517

Bennett, L. W., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2). https://doi.org/10.1177/0267323118760317

Borden, S. L., & Tew, C. (2007). The role of journalist and the performance of journalism: Ethical lessons from “fake” news (seriously). *Journal of Mass Media Ethics*, *22*(4), 300–314. https://doi.org/10.1080/08900520701583586

Boukes, M., Boomgaarden, H. G., Moorman, M., & De Vreese, C. H. (2014). News with an attitude: Assessing the mechanisms underlying the effects of opinionated news. *Mass Communication and Society*, *17*(3), 354–378. https://doi.org/10.1080/15205436.2014.891136

Boukes, M., Boomgaarden, H. G., Moorman, M., & De Vreese, C. H. (2015). At odds: Laughing and thinking? The appreciation, processing, and persuasiveness of political satire. *Journal of Communication*, *65*(5), 721–744. https://doi.org/10.1111/jcom.12173

Boukes, M., & Hameleers, M. (2020). Shattering populists’ rhetoric with satire at elections times: The effect of humorously holding populists accountable for their lack of solutions. *Journal of Communication*, *70*(4), 574–597. https://doi.org/10.1093/joc/jqaa020

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. https://doi.org/10.1177/0956797617714579

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*(1), 20–33. https://doi.org/10.1037/0022-3514.44.1.20

Diddi, A., & LaRose, R. (2006). Getting hooked on news. *Journal of Broadcasting & Electronic Media*, *50*(2), 193–210. https://doi.org/10.1207/s15506878jobem5002_2

Erlich, A., & Garner, C. (2021). Is pro-Kremlin disinformation effective? Evidence from Ukraine. *The International Journal of Press/Politics*. https://doi.org/10.1177/19401612211045221

Feldman, L. (2007). The news about comedy: Young audiences, the daily show, and evolving notions of journalism. *Journalism*, *8*(4), 406–427. https://doi.org/10.1177/1464884907078655

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fox, J. R., Koloen, G., & Sahin, V. (2007). No joke: A comparison of substance in the daily show with presidential election campaign. *Journal of Broadcasting & Electronic Media*, *51*(2), 213–227. https://doi.org/10.1080/08838150701304621

Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication*, *37*(2), 145–156. https://doi.org/10.1080/10584609.2020.1723755

Freire, A. (2015). Left–right ideology as a dimension of identification and of competition. *Journal of Political Ideologies*, *20*(1), 43–68. https://doi.org/10.1080/13569317.2015.991493

Gaines, E. (2007). The narrative semiotics of the daily show. *Semiotica*, *166*, 81–96. https://doi.org/10.1515/SEM.2007.053

Gramlich, J. (2020). *What the data says (and doesn’t say) about crime in the United States*. Retrieved from https://www.pewresearch.org/fact-tank/2020/11/20/facts-about-crime-in-the-u-s/

Hameleers, M., & Van der Meer, T. (2020). Misinformation and polarization in a high-choice media environment: How effective are political fact-checkers? *Communication Research*, *47*(2), 227–250. https://doi.org/10.1177/0093650218819671

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 149–152. https://doi.org/10.1089/cyber.2021.29208.jth

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach’s alpha for estimating reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis* (3rd ed.). The Guilford Press.

Hill, M. R. (2013). Developing a normative approach to political satire: A critical perspective. *International Journal of Communication*, *7*, 324–337. https://ijoc.org/index.php/ijoc/article/view/1934

Holbert, R. (2013). Breaking boundaries: Developing a normative approach to political satire: An empirical perspective. *International Journal Of Communication*, *7*, 1–19. https://ijoc.org/index.php/ijoc/article/view/1933.

Humprecht, E. (2019). Where 'fake news' flourishes: a comparison across four Western democracies. Information, *Communication & Society*, *22*(13), 1973–1988. https://doi.org/10.1080/1369118X.2018.1474241

Jones, J. P., & Baym, G. (2010). A dialogue on satire news and the crisis of truth in postmodern political television. *Journal of Communication Inquiry*, *34*(3), 278–294. https://doi.org/10.1177/0196859910373654

Kim, Y. M., & Vishak, J. (2008). Just laugh! You don't need to remember. *Journal of Communication*, *58*(2), 338–360. https://doi.org/10.1111/j.1460-2466.2008.00388.x

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2017). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research*, *47*(1), 104–124. https://doi.org/10.1177/0093650217719596

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

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

Levine, T. R. (2014). Truth-default theory (TDT) a theory of human deception and deception detection. *Journal of Language and Social Psychology*, *33*(4), 378–392. https://doi.org/10.1177/0261927X14535916

Marchi, R. (2012). With Facebook, blogs, and fake news, teens reject journalistic "objectivity". *Journal of Communication Inquiry*, *36*(3), 246–262. https://doi.org/10.1177/0196859912458700

Meddaugh, P. M. (2010). Bakhtin, Colbert, and the center of discourse: Is there no "truthiness" in humor? *Critical Studies in Media Communication*, *27*(4), 376–390. https://doi.org/10.1080/15295030903583606

Miron, A. M., & Brehm, J. W. (2006). Reactance theory-40 years later. *Zeitschrift für Sozialpsychologie*, *37*(1), 9–18. https://doi.org/10.1024/0044-3514.37.1.9

Nabi, R. L., Moyer-Gusé, E., & Byrne, S. (2007). All joking aside: A serious investigation into the persuasive effect of funny social issue messages. *Communication Monographs*, *74*(1), 29–54. https://doi.org/10.1080/03637750701196896

Nyhan, B., Porter, E., Reifler, J., & Wood, T. (2020). Taking fact-checks literally but not seriously? *Political Behavior*. https://doi.org/10.1007/s11109-019-09528-x

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Ödmark, S. (2021). Making news funny: Differences in news framing between journalists and comedians. *Journalism*, *22*(6), 1540–1557. https://doi.org/10.1177/1464884918820432

Paletz, D. L. (1990). Political humor and authority: From support to subversion. *International Political Science Review*, *11*(4), 483–493. https://doi.org/10.1177/019251219001100406

Richmond, J. C., & Porpora, D. V. (2019). Entertainment politics as a modernist project in a Baudrillard world. *Communication Theory*, *29*(4), 421–440. https://doi.org/10.1093/ct/qty036

Schaewitz, L., Kluck, J. P., Klösters, L., & Krämer, N. C. (2020). When is disinformation (in) credible? Experimental findings on message characteristics and individual differences. *Mass Communication and Society*, *23*(4), 484–509. https://doi.org/10.1080/15205436.2020.1716983

Stroud, N. J., & Muddiman, A. (2013). Selective exposure, tolerance, and satirical news. *International Journal of Public Opinion Research*, *25*(3), 271–290. https://doi.org/10.1093/ijpor/edt013

Strömbäck, J. (2005). In search of a standard: Four models of democracy and their normative implications for journalism. *Journalism Studies*, *6*(3), 331–345. https://doi.org/10.1080/14616700500131950

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, *33*(3), 460–480. https://doi.org/10.1080/10584609.2015.1102187

Vraga, E. K., & Bode, L. (2020). Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. *Political Communication*, *37*(1), 136–144. https://doi.org/10.1080/10584609.2020.1716500

Vraga, E. K., Kim, S. C., & Cook, J. (2019). Testing logic-based and humor-based corrections for science, health, and political misinformation on social media. *Journal of Broadcasting & Electronic Media*, *63*(3), 393–414. https://doi.org/10.1080/08838151.2019.1653102

Waisanen, D. J. (2009). A citizen's guides to democracy inaction: Jon Stewart and Stephen Colbert's comic rhetorical criticism. *Southern Communication Journal*, *74*(2), 119–140. https://doi.org/10.1080/10417940802428212

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research*, *47*(2), 155–177. https://doi.org/10.1177/0093650219854600

Wardle, C. (2017). Fake news. It's complicated. *First Draft*. https://medium.com/1st-draft/fake-news-its-complicated.

Warner, B. R., Jennings, F. J., Bramlett, J. C., Coker, C. R., Reed, J. L., & Bolton, J. P. (2018). A multimedia analysis of persuasion in the 2016 presidential election. *Mass Communication and Society*, *21*(6), 720–741. https://doi.org/10.1080/15205436.2018.1472283

Warner, J. (2007). Political culture jamming: The dissident humor of "The daily show with Jon Stewart". *Popular Communication*, *5*(1), 17–36. https://doi.org/10.1080/15405700709336783

Winter, S., Metzger, M. J., & Flanagin, A. J. (2016). Selective use of news cues. *Journal of Communication*, *66*(4), 669–693. https://doi.org/10.1111/jcom.12241

Wood, T., & Porter, E. (2018). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *41*(1), 135–163. https://doi.org/10.1007/s11109-018-9443-y

Young, D. G. (2008). The privileged role of the late-night joke: Exploring humor's role in disrupting argument scrutiny. *Media Psychology*, *11*(1), 119–142. https://doi.org/10.1080/15213260701837073

Young, D. G., & Hoffman, L. (2012). Acquisition of current-events knowledge from political satire programming: An experimental approach. *Atlantic Journal of Communication*, *20*(5), 290–304. https://doi.org/10.1080/15456870.2012.728121

Young, D. G., Jamieson, K. H., Poulsen, S., & Goldring, A. (2018). Fact-checking effectiveness as a function of format and tone: Evaluating FactCheck.org and FlackCheck.org. *Journalism & Mass Communication Quarterly*, *95*(1), 49–75. https://doi.org/10.1177/1077699017710453