



UvA-DARE (Digital Academic Repository)

Separating truth from lies: comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands

Hameleers, M.

DOI

[10.1080/1369118X.2020.1764603](https://doi.org/10.1080/1369118X.2020.1764603)

Publication date

2022

Document Version

Final published version

Published in

Information, Communication & Society

License

CC BY-NC-ND

[Link to publication](#)

Citation for published version (APA):

Hameleers, M. (2022). Separating truth from lies: comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication & Society*, 25(1), 110-126.
<https://doi.org/10.1080/1369118X.2020.1764603>

General rights

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

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Separating truth from lies: comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands

Michael Hameleers

Political Communication, Amsterdam School of Communication Research (ASCoR), Amsterdam, The Netherlands

ABSTRACT

Although previous research has offered important insights into the consequences of mis- and disinformation and the effectiveness of corrective information, we know markedly less about how different types of corrective information – news media literacy interventions and fact-checkers – can be combined to counter different forms of misinformation. Against this backdrop, this paper reports on experiments in the US and the Netherlands ($N=1,091$) that exposed people to evidence-based or fact-free anti-immigration misinformation, fact-checkers and/or a media literacy intervention. The main findings indicate that evidence-based misinformation is seen as more accurate than fact-free misinformation, and the combination of news media literacy interventions and fact-checkers is most effective in lowering issue agreement and perceived accuracy of misinformation across countries. These findings have important implications for journalism practice and policy makers that aim to combat mis- and disinformation.

ARTICLE HISTORY

Received 2 October 2019
Accepted 29 April 2020

KEYWORDS

Corrective information; disinformation; fact-checkers; misinformation; news media literacy

Mis- and disinformation have been regarded as key threats to deliberative democracy (e.g., Bennett & Livingston, 2018; Humprecht, 2018; Van Aelst et al., 2017). As a consequence of the uncontrolled spread of mis- and disinformation, citizens may become uncertain about the veracity of the basic facts they need as input to make informed political decisions (e.g., Van Aelst et al., 2017). Despite growing concerns on the veracity and truthfulness of political information, extant literature indicates that factual misperceptions can be corrected (e.g., Chan et al., 2017; Clayton et al., 2019; Nyhan et al., 2019; Walter & Murphy, 2018; Wood & Porter, 2018). In general, two types of corrective information can be distinguished: (1) news media literacy interventions or forewarnings (e.g., Clayton et al., 2019; Cook et al., 2017; Tully et al., 2020) and (2) fact-checkers that verify the claims made in (political) communication.

Recent empirical research has mainly focused on the latter approach (e.g., Chan et al., 2017; Nyhan et al., 2019; Wood & Porter, 2018). Although fact-checkers may help to correct

CONTACT Michael Hameleers  m.hameleers@uva.nl  Nieuwe Achtergracht 166, 1018 WV Amsterdam
 Supplemental data for this article can be accessed <https://doi.org/10.1080/1369118X.2020.1764603>

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

misinformation, people show a tendency to avoid fact-checkers that do not align with their perceptual screens (Hameleers & Van der Meer, 2019). Although media literacy interventions may offer a viable alternative journalistic tool to prevent the negative consequences of mis- and disinformation by arming media consumers with knowledge to resist persuasion by communicative untruthfulness, there has been little empirical research on the effectiveness of such forms of corrective information (but see e.g., Cook et al., 2017 and Tully et al., 2020). To date, we know even less about how both types of corrective information may be effective when they are *combined* as an integrated refutation strategy in journalism practice (but see Clayton et al., 2019 and Vraga et al., 2020). The findings of Clayton et al. (2019) indicate that warnings and corrections (false flags) can be effective as they lower the perceived accuracy of false information. Yet, both types of corrective information may have limitations: Warnings have weaker effects than fact-checks and also lower the accuracy of true information (Clayton et al., 2019). Fact-checkers, in turn, only respond to a fraction of all false information, and may have less strong effects under some conditions, although the backfire effect has been disputed in recent work (Nyhan et al., 2019).

This study aims to advance research on mis- and disinformation and corrective information by conducting experiments in the US and the Netherlands in which the effects of (1) news media literacy interventions (2) fact-checkers and a (3) *combination* of pre- and debunking corrections are compared to present new insights into the effectivity of journalist tools to refute communicative untruthfulness. As a key contribution to the literatures on misinformation and corrective information, this study is among the first contributions to experimentally compare the effectiveness of news media literacy interventions and fact-checkers in different settings where misinformation thrives: the US and Europe (Humprecht, 2018).

How mis- and disinformation can result in misperceptions

Misinformation can be defined as information that is not based on empirical evidence and/or expert opinion – it is thus objectively incorrect and empirically falsifiable (e.g., Humprecht, 2018; Nyhan & Reifler, 2010; Tandoc et al., 2018). Although misinformation is spread without the intention to mislead the electorate, empirical evidence has shown that misinformation can affect people's cognitions and attitudes (e.g., Thorson, 2016; Wood & Porter, 2018). Different from misinformation, disinformation entails the dissemination of information that is *deliberately* manipulated, fabricated or placed in a different context in order to achieve a political goal (e.g., Marwick & Lewis, 2017; Wardle, 2017). Agents of disinformation, such as radical right-wing politicians, may spread falsehoods to achieve political success, shift blame to political opponents for failures, or even to disrupt the political and societal order (e.g., Bennett & Livingston, 2018; Marwick & Lewis, 2017).

We distinguish different types of argumentation or framing that can be used in mis- and disinformation. Mis- or disinformation may, first of all, lack empirical evidence or expert knowledge (Nyhan & Reifler, 2010). Statements in misinformation may thus be classified by the lack of facticity – or at least the absence of references to authentic sources of facts and verification (Tandoc et al., 2018). This type of misinformation reflects news reporting that relies on popular exemplars and experiences of ordinary people to illustrate cases and wider issues rather than offering (hard) empirical evidence and expert knowledge (e.g., Lefevere et al., 2011).

An alternative form of misinformation that profits from the legitimacy of journalistic formats using references to (fake) sources and facts may be perceived as more accurate

than fact-free misinformation that circumvents references to verified knowledge. By profiting from the legitimacy of mainstream journalism, lies can be sold more efficiently using a layer of facticity (Balod & Hameleers, 2019). Especially on topics such as immigration and crime-rates, facts and statistics may be more effective than experiences and exemplars. We introduce the following hypothesis on the differential effects of the type of argumentation used in mis- or disinformation: Mis- or disinformation that contains references to verified knowledge and expert sources is perceived as more accurate (H1a) and yields higher levels of issue agreement (H1b) than mis- or disinformation that relies on people's experiences without references to facticity.

The resonance of misinformation with ideological perceptual screens

In line with the psychological mechanisms of motivated reasoning (Taber & Lodge, 2006) misinformation may be accepted as a consequence of the persistence of confirmation biases and directional motivated reasoning (e.g., Nyhan et al., 2019). Hence, when people's perceptual screens align with information, they are more likely to accept this information to remain at cognitive consonance, irrespective of its veracity (Hameleers & Van der Meer, 2019). For misinformation on immigration, we can thus expect that people with congruent issue attitudes are more likely to perceive misinformation as accurate compared to people with incongruent issue attitudes. We therefore hypothesize that misinformation that is congruent with people's pre-existing immigration attitudes is more likely to be perceived as accurate (H2a) and yields higher levels of issue agreement (H2b) compared to misinformation that is incongruent with pre-existing views.

Finally, it can be expected that people-centric misinformation resonates more among participants with anti-immigration perceptions than misinformation that relies on (false) facts and expert knowledge. Anti-immigration beliefs correspond with a right-wing populist perspective, which has been associated with lower levels of trust in elite and expert sources. People with right-wing populist views are more likely to be attracted to media formats that emphasize people-centrism compared to information that relies on expert knowledge (Hameleers et al., 2017). We therefore expect that fact-free or people-centric compared to expert-based misinformation is perceived as more accurate (H2c) and yields higher levels of issue agreement (H2d) among participants with more pronounced anti-immigration attitudes.

The effects of corrective information: media literacy interventions and fact-checkers

Although recent empirical research has focused on the effects of retractions in the form of fact-checkers (e.g., Cobb et al., 2013; Hameleers & Van der Meer, 2019; Nyhan et al., 2019; Wood & Porter, 2018) – we know markedly little about the effects of news media literacy interventions (Cook et al., 2017 and Tully et al., 2020). Even more important, there has been little research that systematically *compared* the effects of both types of corrective information (but see Clayton et al., 2019 and Vraga et al., 2020). In this setting, this paper explores the effects of (1) news media literacy messages; (2) fact-checkers and (3) the *combination* of both types of corrective information on misinformation's accuracy and issue agreement with (rebutted) false information.

The effects of news media literacy interventions

News media literacy can be regarded as the skills and knowledge that news consumers need to navigate their information environment in a mindful and critical way (Aufderheide, 1993; Ashley et al., 2017; Tully et al., 2020). When news consumers are literate, this means that they understand how (political) information is produced, consumed, and how personal biases and existing beliefs may play a role in how news is interpreted. News media literacy (NML) approaches can help to stimulate more critical skills, for example by warning people about the negative impact of misleading information (Clayton et al., 2019; Tully et al., 2020). In general, NML interventions can enhance critical media skills by (1) informing people about how news content is produced and consumed; (2) enhancing knowledge about the impact that information may have on society, politics, and individual media consumers, and (3) revealing the disconnect between the mediated reality and the external reality (e.g., Jeong et al., 2012). Especially in the digital age, it may be important that (young) citizens know how to assess the quality, truthfulness and honesty of information (e.g., Kahne et al., 2012).

Applied to misinformation, media literacy messages should inform people on how to recognize false information, inform people about the consequences of misleading information, and learn people how to distinguish truthful from false information (also see Clayton et al., 2019; Tully et al., 2020). In this paper, we base the design of a NML intervention on these theoretical premises. Specifically, we designed a message that (1) warns citizens about the existence of misleading information; (2) informs how false information may be detected by looking at the source and type of evidence and argumentation offered and (3) distinguish the external reality from the (biased) media depiction of reality. This also corresponds to the approach of Jones-Jang et al. (2019): Media literacy skills that help people to find reliable and accurate sources of factual information increase the likelihood that people can detect false information. Finally, our approach is in line with Clayton et al.'s (2019) design of a warning message that provided advice on how to recognize false information, which mimics actual applications of misinformation warnings used by Facebook.

Together, based on recent research conducted by Cook et al. (2017), Clayton et al. (2019) and Tully et al. (2020) on the design of potentially effective news media literacy interventions as well as real-life examples, we explore the effectiveness of news media literacy interventions that offer a forewarning on the techniques of misinformation, as well as practical tips on how to avoid misleading content. These interventions should be effective as they (1) foreground recommendations on how to recognize faulty lines of argumentation in communicative untruthfulness (i.e., no reliance on factual information and expert opinion) and (2) explicate the motives of agents of disinformation, and (3) explain the techniques underlying the (viral) spread of mis- and disinformation. Against this backdrop, we hypothesize: Exposure to a media literacy message leads to lower levels of perceived accuracy of (H3a), and agreement with (H3b), the claims made in mis- and disinformation compared to the absence of a news media literacy message.

The effectiveness of fact-checkers

A growing body of research indicates that exposure to fact-checkers may lower the perceived accuracy of, and agreement with, communicative untruthfulness (e.g., Chan

et al., 2017; Hameleers & Van der Meer, 2019; Nyhan et al., 2019; Walter & Murphy, 2018; Wood & Porter, 2018). Fact-checkers' reliance on short, factual arguments that directly respond to inaccurate statements in mis- and disinformation may be effective. Compared to news media literacy interventions, fact-checkers present a more direct and less abstract counterargument to (political) claims made in news stories. Fact-checkers also offer a direct and clear verdict of the scope of communicative untruthfulness (i.e., the article is mostly false). Yet, some studies on the effectiveness of political fact-checking have indicated that a backfiring effect may occur (e.g., Nyhan & Reifler, 2010; Thorson, 2016). This means that when strong partisans are exposed to a fact-checker that attacks their prior attitudes or ideological identification, they may strengthen rather than lower their partisan beliefs. Although some recent studies found no evidence for this backfiring effect (e.g., Nyhan et al., 2019; Wood & Porter, 2018), fact-checking may be less effective among people that agree with the claims made in misinformation (Hameleers & Van der Meer, 2019).

In addition, fact-checkers are found to be least effective in correcting political misinformation (Walter & Murphy, 2018). Another potential 'risk' of fact-checking is that it has to correct or overrule cognitive or affective associations that are already stored in news consumers' minds. Hence, fact-checkers typically do not directly follow misinformation, which means that they have to correct misperceptions that are already part of people's schemata. This may reduce the real-world impact of fact-checking as corrective information as opposed to interventions that aim to prevent the cultivation of misperceptions.

Even though partisans and issue publics are motivated to defend their (party) identification and ideological beliefs, and even if cognitively stored associations may be hard to overrule, factual corrections have been found to even impact the *factual* perceptions of stronger partisans (Nyhan et al., 2019). Clayton et al. (2019) offer first evidence that corrective information that flags false information has a stronger impact on the perceived accuracy of information than general warnings. Yet, the flag used to refute misinformation needs to explicitly note that the article is false. Against this backdrop, we hypothesize that exposure to a fact-checker leads to lower levels of perceived accuracy of (H4a) and agreement with (H4b) the claims made in mis- and disinformation compared to the absence of a fact-checker. The same pattern should apply to information that confirms misinformation: exposure to verification should *increase* perceived accuracy and issue agreement.

As both types of corrective information may have clear merits as well as limitations, we argue that the *combined* use of media literacy interventions and fact-checkers as journalist tools may provide the most effective strategy to refute mis- and disinformation. Whereas Clayton et al. (2019) argue that general warnings are less effective than post-hoc corrections, and that general warnings do not augment the effects of simple flags, the combination of a news media literacy message and a comprehensive fact-checker may be most effective. The *combination* of a forewarning or inoculation message that helps citizens to recognize mis- and disinformation and a fact-checker that actually confirms the communicative untruthfulness that people are exposed to may offer the clearest and most powerful retraction of dishonest communication. Yet, it should be noted that Vraga et al. (2020) found that exposure to a news literacy message did not enhance the effect of an expert correction, at least when looking at factual misperceptions as dependent variable. The question remains if this finding holds for the effects of different types of

corrections on different outcome variables (accuracy and issue agreement) in the US and the Netherlands.

In this paper, next to assessing the isolated and independent effect of fact-checkers and media literacy interventions, we explicitly estimate the effectiveness of a news media literacy intervention preceding exposure to misinformation and a fact-checker responding to misinformation. We hypothesize that exposure to *both* a media literacy intervention and a fact-checker may result in lower levels of perceived accuracy of (H5a) and agreement with (H5b) communicative untruthfulness than exposure to *only* a fact-checker or media literacy intervention. We thus expect a combined approach to be more effective than a single approach to correcting misinformation.

Although most research on misinformation and corrections is conducted in the US, misinformation is relevant to consider in different settings. As indicated by the Reuters News Report (2018), concerns on mis- and disinformation are much more salient in the US than in the Netherlands. The content of misinformation also differs across national settings: US-based misinformation relies on more partisan framing, whereas European misinformation, at least in German-speaking countries, is mostly about immigration (Humprecht, 2018). In these different settings, this paper aims to explore how universal the effects of misinformation on immigration and crime rates are in the Netherlands and the US. Here, it should be noted that we do not expect to find differences across the different national settings. Rather, the aim to conduct the study in two different countries is to validate and generalize findings in settings that have dealt with misinformation and corrective information in different ways: Are the effects of misinformation, fact-checkers, and media literacy messages similar across different national settings?

Method

Design

In both countries, participants were randomly assigned to one of the between-subjects experimental conditions. More specifically, the design concerns a 2 (Pre-bunking media literacy message: absent versus present) x 2 (Anti-immigration mis- or disinformation: evidence-based versus fact-free) x 3 (fact-checker: rebuttal versus confirmation versus absence). The group sizes were equal in both countries. If participants were exposed to a pre-bunking message, this media literacy intervention was not placed directly in front of exposure to misinformation, but was followed by a question block (the same questions were asked to participants in the no-pre-bunker condition). The fact-checker followed exposure to misinformation – with a short distractor in between (an instruction to think about the article – and a forced exposure block of 20 s).

Sample

A representative sample of U.S. and Dutch participants was collected by an international research company. Eligible participants were 18 years or older ($M = 43.54$, $SD = 14.14$). The total number of completes was 1,091 (Netherlands: 546 US: 545). As the sample size per condition is relatively small, we applied bootstrapping techniques in our analyses. Power analysis confirms that this sample size is sufficient to detect statistically significant differences between conditions.

The completion rate was 87.1%. Of all participants, 45.9% was female, 22.9% was lower educated, and 23.5% higher educated. 40.7% self-identified with a left-wing ideological position, and 46.6% identified themselves as right-wing. Regarding issue positions toward the topic of misinformation in this experiment, immigration, 38.7% opposed immigration, 37.2% supported immigration, and 24.0% did not support or oppose immigration. There were no noteworthy differences in sample distributions across the two countries.

Independent variables

Mis- or disinformation. We varied the type of argumentation used in misinformation: evidence-based or factual coverage versus anti-experts and people-centric coverage resonating with opinions and experiences (the stimuli are included in Appendix A). In all conditions, the topic of the misinformation message was increasing crime rates. Although the real numbers show a consistent decline in all crime rates in the US and the Netherlands, the fake news story argued the opposite: violent crimes were said to increase – and this was connected to the threat coming from immigrants. This development was presented as a threat to the native people in both countries. The evidence-based news story quoted a fake expert (a professor) and referred to non-existing statistics of the national statistics bureau to provide false evidence for the negative development of the crime rate. Moreover, a non-existing research project was invented as a source for the predicted developments.

In the anti-experts and people-centric misinformation condition, references to expert sources and empirical evidence were excluded from the narrative. The same developments (violent crime rates) and causes (immigrants) were presented, but the ordinary people and public opinion were used as the source of knowledge – which corresponds to people-centric news coverage and formats that are popular in today's media environments. The article quoted a panel of 'ordinary citizens' to contextualize the developments of the increasing violent crime rate caused by immigrants. The stimuli were very similar across both countries – only the country names were changed (overall percentage-wise statistics, sources and references were matched and pre-tested on accuracy and similarity).

Media literacy intervention. We based our media literacy intervention on existing formats used in both the US and the Netherlands as tools to combat the consequences of mis- and disinformation – hereby maximizing the external validity of our approach. Based on theory on the persuasiveness of forewarnings, we further ensured that the intervention was practical, concrete, short, and stimulated efficacy beliefs by presenting participants with concrete, easy to use tools to avert the threat (see Appendix A for the media literacy intervention). The media literacy intervention emphasized three key recommendations for media consumers to recognize misinformation: (1) check the source of the message and the sources quoted in the message; (2) search for the facts, and assess if these facts are actually seen as accurate in light of the developments presented and (3) assess whether the argumentation makes sense: are the consequences and causes logically connected?

Fact-checkers. Again, the manipulations were based on existing journalistic tools that are used to combat mis- and disinformation, such as PolitiFact.com and factcheck.org and Dutch alternatives, such as *Nieuwscheckers*. Similar to existing fact-checkers, an explicit recommendation on the veracity of the information was given: completely true in the confirmation condition and completely false in the refutation condition. In the refuting fact-check condition, all claims made in the fake news story were refuted based on real empirical evidence, for example from research on crime rate developments by Oxford

University. The causal connection between increasing crime rates and dangerous immigrants was also refuted based on real empirical evidence and expert opinion.

The confirming fact-checker presented the opposite view. The same sources and references to empirical evidence were used, but manipulated to support the claims made in the false news article: the research project by Oxford University was said to support the development of increasing crime rates caused by the influx of dangerous migrants.

Dependent variable: perceived accuracy

The perceived accuracy of the (fictional) news items was measured with five items measured on a 7-point (disagree–agree) scale: (1) *The news item is truthful*, (2) *The news item can be regarded as Fake News*, (3) *The news item is not accurate*, (4) *The news item deviates from reality*, (5) *the news item does not cover the facts as they happened* (Cronbach's $\alpha = .849$, $M = 3.94$, $SD = 1.26$). Items in the scale were recoded to reflect perceptions of accuracy (most individual items tapped negative perceptions toward the accuracy of the article, which were reverse-coded to measure accuracy).

Dependent variable: issue agreement

To measure issue agreement after exposure to different forms of misinformation and corrective efforts, a nine-item scale tapped into participants' agreement with the anti-immigration statements expressed in the article. The items include: (1) *Our safety is threatened by migrants*, (2) *The crime rate in our country is worsening because of the elite's failing policies*, (3) *We need to have stronger background checks on migrants*, (4) *We need to close our borders for migrants*, (5) *We need better solutions to deal with the influx of migrants*; (6) *Migrants are responsible for violent crimes*; (7) *The political elite's policies on migration are failing*; (8) *Politicians need to do more to deal with the issue of illegal immigration*; (9) *Political elites are responsible for problems related to migration* (Cronbach's $\alpha = .929$, $M = 4.68$, $SD = 1.45$). This measure was different from the pre-exposure measure of anti-immigration beliefs, which tapped more general perceptions toward immigrants.

Moderator: anti-immigration beliefs

Prior to exposure, anti-immigration attitudes were measured with a battery of five items measured on a 7-point (disagree – agree) scale (i.e., 'migrants pose a threat to our safety' and 'migrants are more inclined to commit violent crimes than native people' (Cronbach's $\alpha = .794$, $M = 4.22$, $SD = 1.45$). The items were related to the fictional news story that connected immigrants to violent crimes. Yet, we made sure that these items were formulated in a different way than the post-test dependent variable of issue agreement.

Pilot test and manipulation checks

The stimuli were pilot tested among a varied convenience sample ($N = 56$). In this pilot test, participants were asked to rate the accuracy of the media literacy intervention, the fact-checkers and the misinformation. Overall, the corrective information was found to be very accurate ($M = 5.00$, $SD = 1.63$) and the manipulated information was found to

be similar to the coverage people encountered in their daily media environment ($M = 5.96$, $SD = 1.57$). In the main study, manipulation check items were included at the end of the survey. First of all, the evidence-based version of misinformation was found to be significantly and substantially more factual than the fact-free message ($\Delta M = 1.29$, $\Delta SE = .24$, $t = 5.33$, $p < .001$). Participants exposed to a refuting fact-checker were significantly more likely to recognize the corrective information as counter-arguing the misinformation compared to participants that did not see such a factual rebuttal ($\Delta M = 1.63$, $\Delta SE = .17$, $t = 9.58$, $p < .001$). The manipulation for the confirmatory fact-checker was also successful: participants that were exposed to a confirmation were more likely to believe that the fact-checker confirmed the issue positions of the article compared to participants that did not see a conformation ($\Delta M = 1.57$, $\Delta SE = .17$, $t = 9.29$, $p < .001$). Finally, the media literacy training manipulation succeeded. Participants exposed to a pre-bunker were overall very likely to correctly identify the three key recommendations emphasized in the media literacy intervention (87.3% correct).

Results

The persuasiveness of fact-free and evidence-based (Mis)information

Regarding the overall perceived accuracy of fact-free versus evidence-based misinformation, the findings depicted in Table 1 (Model I) indicate that untrue communication that circumvents evidence and factual coverage is perceived as significantly *less* accurate than evidence-based coverage of immigration news. There is no significant interaction effect between exposure to misinformation without factual evidence and the national setting (Table 1, model III), albeit misinformation without factual references is perceived as less accurate in the US than the Netherlands (for illustrative purposes, we have included country-specific estimates in Figure 1 of the supplemental material). The effect of evidence-type is similar for the effects of misinformation on issue agreement (Table 2, Model I). Hence, H1b also finds support: Exposure to misinformation that circumvents factual information results in *lower* levels of issue agreement compared to misinformation that relies on evidence-based coverage. There is again no significant interaction effect between country and type of misinformation on issue agreement (Table 2, Model III).

We further expected that misinformation that aligns with prior attitudes is perceived as more accurate than incongruent misinformation (H2a). First of all, the findings indicate

Table 1. OLS- Regression model predicting the perceived accuracy of misinformation.

	Model I (n = 1,091)			Model II (n = 1,091)			Model III (n = 1,091)		
	B	SE	β	B	SE	β	B	SE	β
(Constant)	4.02	.16		3.35	.16		3.49	.24	
Fact-free communication	-.14	.07	-.06*	-.12	.08	-.05	-.35	.31	-.14
Country (Netherlands)				.13	.08	.05	-.01	.12	-.01
Anti-immigration				.24	.03	.28***	.14	.05	.16**
Fact-free \times country							.25	.16	.09
Fact-free \times anti-immigration							.16	.06	.30**
Adjusted R^2		.002			.070			.080	
F		3.36			21.56***			14.44***	
F for change in R^2					27.54***			4.66**	

Note. Two-tailed tests. Unstandardized (B) and standardized (β) regression weights. Bootstrapping was employed.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 2. OLS- Regression model predicting levels of issue agreement with misinformation.

	Model I (n = 1,091)			Model II (n = 1,091)			Model III (n = 1,091)		
	B	SE	β	B	SE	β	B	SE	β
(Constant)	4.80	.07		1.34	.14		1.47	.21	
Fact-free communication	-.20	.09	-.07*	-.13	.06	-.04	-.33	.26	-.11
Country (Netherlands)				.24	.07	.08***	.14	.10	.05
Anti-immigration				.62	.03	.62***	.51	.04	.51***
Fact-free \times country (NL)							.18	.13	.05
Fact-free \times anti-immigration							.18	.05	.30***
Adjusted R ²		.004			.496			.504	
F		5.29*			269.41**			159.32***	
F for change in R ²					355.73***			6.79***	

Note. Two-tailed tests. Unstandardized (B) and standardized (β) regression weights. Bootstrapping was employed. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

that participants' prior issue positions related to immigration positively and significantly correspond to the perceived accuracy of misinformation (Table 1, Model II). People thus are most likely to perceive misinformation as accurate when it reassures their existing beliefs. In support of H2b, issue publics are also most likely to agree with communicative untruthfulness. Misinformation is thus perceived as more credible, and yields higher levels of issue agreement, when participants' prior attitudes align with the message.

The interaction effect between exposure to fact-free misinformation and anti-immigration attitudes is positive and significant (Table 1, Model III), which indicates that participants with congruent prior attitudes are more likely to accept fact-free coverage than participants with incongruent priors. The same effects were found for issue agreement (Table 2, Model III). This is in line with H2c and H2d: the more pronounced participants' prior anti-immigration attitudes, the more likely they are to be persuaded by fact-free compared to evidence-based misinformation.

The effects of media literacy interventions and fact-checkers

In Tables 3 and 4, we compared the effects of different corrective attempts in response to untrue communication. The findings show that exposure to a media literacy message significantly lowers the perceived accuracy of misinformation (Table 3, Model II). This effect is similar for evidence-based and fact-free misinformation (the two-way interaction effect between a media literacy message and fact-free misinformation is non-significant: $B = -.42$, $SE = .28$, $p = n.s.$). These findings support H3a: exposure to a media literacy message lowers the perceived accuracy of communicative untruthfulness. However, exposure to a media literacy message does not result in lower levels of agreement with the statements made in misinformation (Table 4, Model II). Again, there are no significant country differences (see the non-significant interaction effects depicted in Table 3 and 4, Model IV). This means that H3b does not find support in the data: although news media literacy interventions can lower the perceived accuracy of misinformation, they do not decrease the overall levels of agreement with communicative untruthfulness.

Turning to H4a, the results indicate that exposure to a fact-checker debunking misinformation has a positive effect in the desired direction of the correction (see Table 3, Model II). More specifically, when a fact-checker confirms that the message was correct, participants were significantly more likely to perceive the message as accurate compared

Table 3. OLS- Regression model predicting the effects of corrective information on perceived message accuracy.

	Model I (n = 1,091)			Model II (n = 1,091)			Model III (n = 1,091)			Model IV (n = 1,091)		
	B	SE	β	B	SE	β	B	SE	β	B	SE	β
(Constant)	2.98	.12		3.19	.16		2.97	.28		3.02	.29	
Country (Netherlands)	.22	.07	.09**	.20	.07	.08**	.20	.07	.08*	.11	.19	.05
Anti-immigration	.20	.03	.23***	.20	.03	.23***	.25	.06	.29***	.25	.06	.29***
Fact-free communication				-.08	.08	-.03	-.08	.08	-.03	-.11	.08	-.04
Media literacy				-.25	.12	-.08*	.28	.38	.09	.22	.39	.07
Fact-check confirmation				.33	.13	.10**	.14	.39	.04	.08	.39	.02
Fact-check rebuttal				-.45	.13	-.13***	-.32	.39	.09	-.47	.40	-.14
Combined confirmation				.01	.13	.01	.39	.39	.11	.26	.40	.07
Combined rebuttal				-.46	.13	-.13***	.32	.39	.09	.35	.41	.10
Media literacy \times immigration							-.12	.08	-.18	-.13	.08	-.19
FC refute \times immigration							.03	.09	.03	.03	.09	.03
FC confirm \times immigration							.04	.08	.06	.03	.09	.04
Combined confirm \times immigration							-.09	.09	-.11	-.10	.09	-.12
Combined refute \times immigration							-.18	.09	-.24*	-.18	.09	.24*
Media literacy \times country (NL)										.18	.25	.04
FC refute \times country (NL)										-.14	.25	-.03
FC confirm \times country (NL)										.21	.26	.05
Combined confirm \times country (NL)										.31	.26	.07
Combined refute \times country (NL)										-.06	.26	-.01
F		.063			.109			.114			.114	
F for change in R^2		37.63***			17.72***			11.83***			8.80***	
					10.43***			2.24*			.95	

Note. Two-tailed tests. Unstandardized (B) and standardized (β) regression weights. Bootstrapping was employed.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 4. OLS- Regression model predicting the effects of corrective information on issue agreement.

	Model I (n = 1,091)			Model II (n = 1,091)			Model III (n = 1,091)			Model IV (n = 1,091)		
	B	SE	β	B	SE	β	B	SE	β	B	SE	β
(Constant)	1.71	.10		1.89	.25		1.89	.25		1.90	.25	
Country (Netherlands)	.16	.06	.06*	.14	.06	.05*	.14	.06	.05*	.14	.16	.05
Anti-immigration	.69	.02	.69***	.69	.05	.69***	.69	.05	.69***	.69	.05	.69***
Fact-free communication				-.15	.07	-.05*	-.15	.07	-.05*	-.17	.07	-.06
Media literacy				-.11	.11	-.03	-.16	.33	-.04	-.14	.34	-.04
Fact-check confirmation				.16	.11	.04	.71	.33	.19*	.67	.34	.18*
Fact-check rebuttal				-.26	.11	-.07*	-.58	.31	-.15*	-.50	.35	-.13
Combined confirmation				-.10	.11	-.03	.18	.34	.05	.09	.35	.02
Combined rebuttal				-.35	.11	-.09***	-.60	.34	-.15*	-.57	.35	-.14
Media literacy \times immigration							.01	.07	.01	.02	.07	.02
FC refute \times immigration							.08	.08	.09	.08	.08	.09
FC confirm \times immigration							-.13	.07	-.15	-.14	.07	-.15
Combined confirm \times immigration							-.07	.08	-.08	-.07	.08	-.08
Combined refute \times immigration							.06	.08	.07	.06	.08	.07
Media literacy \times country (NL)										-.09	.22	-.02
FC refute \times country (NL)										-.17	.22	-.03
FC confirm \times country (NL)										.13	.22	.03
Combined confirm \times country (NL)										.19	.22	.04
Combined refute \times country (NL)										-.06	.23	-.01
Adjusted R^2		.481			.493			.496			.496	
F		505.75***			133.65***			83.54***			60.49***	
F for change in R^2					5.46***			2.19*			.79	

Note. Two-tailed tests. Unstandardized (B) and standardized (β) regression weights. Bootstrapping was employed.

* $p < 0.05$; *** $p < 0.001$.

to when such a correction was absent. Likewise, when the fact-checker *refutes* the information of the political message on anti-immigration, participants' perceived accuracy is significantly lower than when this correction is absent. There are no significant country-differences, which indicates that corrective information works similarly across both national settings (Table 3, Model IV). Overall, the results thus provide support for H4a: fact-checkers can correct misinformation, and can also confirm the veracity of information when they conclude that the message is correct.

In a similar vein, exposure to a fact-checker that refutes misinformation lowers participants' agreement with anti-immigration misinformation (Table 4, Model II). More specifically, participants that are exposed to a fact-checker are less likely to agree with misinformation than participants that did not see a fact-checker. Again, the interaction effect between exposure to corrective information and country is non-significant (Table 4, Model IV). These findings support H4b: Fact-checkers successfully lower the perceived accuracy of false information, and reduce agreement with the falsehoods communicated.

As can be seen in Table 3 (Model II), the combination of a media literacy and fact-checking message does not have a stronger impact on participants' evaluation of accuracy than mere exposure to a fact-checker that refutes misinformation. This does not support H5a. However, supporting H5b, exposure to a media literacy message and a fact-checker combined has a stronger negative effect on issue agreement than exposure to a fact-checker or media literacy intervention alone (Table 4, Model II). These findings are similar when we look at the different national settings: there is no significant interaction effect between exposure to combined refutations and the national setting (Table 3 and 4, Model IV).

Discussion

As the honesty and veracity of information is at risk in today's post-truth information settings, where different actors intentionally or unintentionally mislead news audiences by spreading accurate information alongside inaccurate or fabricated content, it is crucial to assess how misperceptions resulting from exposure to misinformation can be corrected (also see e.g., Nyhan et al., 2019; Wood & Porter, 2018). Against this backdrop, this paper relied on experiments in the US and the Netherlands to investigate how different forms of misinformation may mislead the electorate, and how misperceptions may be corrected by different journalistic tools: fact-checkers and media literacy interventions.

Our key findings indicate that misinformation that uses fake statistics, experts and evidence is perceived as more accurate than misinformation without factual references. In line with the politics of disinformation, different actors may strategically manipulate or fabricate stories to respond to people's confirmation bias, which for example resonates with the persuasion tactics of radical right-wing leaders (e.g., Bennett & Livingston, 2018; Marwick & Lewis, 2017). At least in the setting of U.S. political communication, when these actors come up with fake facts and sources to give their story more evidential value, their message may be seen as more accurate, and even augment receivers' agreement with the fabricated content.

The findings further indicate that, irrespective of its veracity, misinformation is perceived as more accurate and persuasive when it confirms pre-existing beliefs. Again, this may have far-reaching ramifications for democracy. In a post-truth and high-choice

information ecology, media consumers can select different (ideological) framings of the same issue (Van Aelst et al., 2017). As selective exposure research indicates that congruent information has a higher chance of selection than incongruent information (Knobloch-Westerwick et al., 2017), citizens may select their own biased version of the truth that reassures their prior attitudes, which means that the objective reality becomes subject to interpretation.

These effects indicate that we should be worried about misinformation's impact on society. In the next step, we therefore investigated how communicative untruthfulness may be combated. We specifically compared the effects of two journalistic tools and their interactions: media literacy interventions or inoculation tactics (e.g., Cook et al., 2017; Tully et al., 2020) and fact-checkers (e.g., Hameleers & Van der Meer, 2019; Nyhan & Reifler, 2010; Wood & Porter, 2018). Overall, we found that exposure to a media literacy intervention only has a significant effect on the perceived accuracy of misinformation, and not on issue agreement. Hence, media consumers' level of agreement with misinformation cannot be corrected effectively by relying on media literacy messages alone.

Our findings do, however, indicate that the *combination* of a media literacy intervention and a fact-checker that refutes falsehoods is most effective. Such integrative interventions help to correct communicative untruthfulness in both countries, and it has an effect on both misperceptions (issue agreement) and the accuracy of misinformation. Our findings are in line with Clayton et al.'s (2019) conclusions: warnings about misinformation and corrections may both be effective tools to correct misinformation. However, the findings are not in line with Vraga et al.'s (2020) experiments, which demonstrate that news media literacy messages do not enhance the effects of expert sources of fact-checking. The placement of the different types of corrections as well as its arguments and formats may explain these differences, which makes it relevant for future research to experiment with formats, placement and argument-types of corrective information.

Both types of corrections come with risks: media literacy messages or warnings have weaker effects than corrections (Walter & Murphy, 2018), but fact-checkers may have a hard time correcting existing schemata and stored cognitions if the correction is delayed (which is very likely in the digital communication setting characterized by information overload and fragmentation). Finally, pre-warning messages have the risk of increasing skepticism and cynicism among news consumers, who may overestimate the presence of false information in their media environment. As a practical implication, it is important to formulate pre-warning messages in a way to induce 'healthy skepticism' rather than inducing distrust in the media. In addition, warnings should formulate recommendations that apply to mis- and disinformation and not to verified journalism. We need further empirical research to show whether media literacy messages also reduce the perceived accuracy of real information, and whether they increase distrust in the media on more general levels.

Our findings do point to a potential negative side-effect of exposure to fact-checkers. Moving beyond existing research that mainly exposed participants to rebuttals of political news (e.g., Hameleers & Van der Meer, 2019), we also assessed how (fake) fact-checkers that verify (mis)information can impact misperceptions. Although such confirmations did not have an impact in the US, Dutch news consumers perceived misinformation significantly and substantially more accurate when it was reinforced by a confirming fact-

checker. This means that we have to place a critical side-note to the practical implications of fact-checkers. Although they may be extremely valuable tools to combat misinformation when in the right hands, communicators with the wrong intentions may profit from the legitimacy and perceived accuracy of fact-checkers and use their format to reinforce disinformation – hereby making falsehoods even more credible by allegedly verifying it with fake evidence. As numerous ‘fake’ fact-checkers are launched in online, alternative media settings (the Swedish fact-check platform *faktiskt.se* is, for example, copied by agents of disinformation), it should be noted that it is important to safeguard the authenticity and independence of fact-checkers.

In line with more recent research that did not replicate a backfire effect of corrective information (e.g., Nyhan et al., 2019; Wood & Porter, 2018), people with prior anti-immigration attitudes did not respond differently to corrective efforts. This means that corrective efforts can help to combat misinformation, even among people that tend to agree with the issue positions emphasized in incorrect or dishonest information. Yet, in real-life information ecologies characterized by high-choice and overload, fact-checkers may be less effective than in an experimental setting. Hence, fact-checkers typically do not directly follow misinformation, and fact-checkers need to be selected in order to have an effect (Hameleers & Van der Meer, 2019). Although it may be relatively easy to counter falsehoods that are not yet part of people’s schemata, misperceptions that persist for a longer period before being corrected need to override or change existing associations.

This paper is not without its limitations. We forced people into exposure to misinformation and corrective information, whereas citizens may avoid misinformation or corrective information in real life, especially when it attacks their prior beliefs (Hameleers & Van der Meer, 2019). Future research should assess the likelihood of selective exposure and avoidance into different forms of (in)congruent misinformation and corrections. Second, we only manipulated misinformation on one highly salient and polarizing topic in both countries – for which prior attitudes play a more decisive role than any other factor manipulated in the experiment. The experimental set-up only allowed us to identify short-term effects, whereas corrective efforts and misinformation are typically more fragmented in today’s high-choice digitized media settings. Future research may rely on multi-wave experiments to tease out the longer term effects of exposure to misinformation, as well as the duration of corrective efforts. In addition, our measure of perceived accuracy points to both perceptions of inaccuracy without intent and items that include the weaponized term Fake News (which may signal intentional deception). Although we could not differentiate two separate dimensions, and found similar effects with or without the Fake News item, we suggest future research to distinguish perceptions of misinformation from disinformation. Finally, we need more research on different types of corrective information: the media literacy intervention designed for this experiment was relatively short – and consisted of one single and relatively short message. Although the type of argumentation was in line with existing approaches to news media literacy (Tully et al., 2020), future research may explore the (longer term) effects of repetition and different formats. As indicated by Jones-Jang et al. (2019), effective media literacy interventions should enhance skills to find online resources that are accurate, reliable, and verified.

Despite these limitations, this study makes an important contribution by demonstrating that misinformation can be seen as accurate and may foster issue agreement, especially when journalistic routines of verified evidence and expert references are used. However,

journalists and governments play an important role in combating the spread of misinformation as they can, at least partially, counter the political consequences of communicative untruthfulness by strengthening their roles of educators, watchdogs, and fact-checkers.

Disclosure statement

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

Notes on contributor

Michael Hameleers (PhD, University of Amsterdam) is Assistant Professor in Political Communication at the Amsterdam School of Communication Research (ASCoR), Amsterdam, The Netherlands. His research interests include populism, framing, (affective) polarization, and the role of social identity in media effects. All authors have agreed to the submission and the article is not currently being considered for publication by any other print or electronic journal.

References

- Ashley, S., Maksl, A., & Craft, S. (2017). News media literacy and political engagement: What's the connection. *Journal of Media Literacy Education*, 9(1), 79–98. <https://doi.org/10.23860/JMLE-2017-9-1-6>
- Aufderheide, P. (1993). *Media Literacy. A Report of the National Leadership Conference on Media Literacy*. Aspen Institute, Communications and Society Program, 1755 Massachusetts Avenue, NW, Suite 501, Washington, D.C.
- Balod, H. S., & Hameleers, M. (2019). Fighting for truth? The role perceptions of Filipino journalists in an era of mis- and disinformation. *Journalism*. <https://doi.org/10.1177/1464884919865109>
- Bennett, L. W., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, <https://doi.org/10.1177/0267323118760317>
- 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>
- Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2019). 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*, 1–23. <https://doi.org/10.1007/s11109-019-09533-0>.
- Cobb, M. D., Nyhan, B., & Reifler, J. (2013). Beliefs don't always persevere: How political figures are punished when positive information about them is discredited. *Political Psychology*, 34(3), 307–326. <https://doi.org/10.1111/j.1467-9221.2012.00935.x>
- Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, 12(5), <https://doi.org/10.1371/journal.pone.0175799>
- Hameleers, M., Bos, L., & de Vreese, C.H. (2017). They did it”: The effects of emotionalized blame attribution in populist communication. *Communication Research*, 44(6), 870–900. <https://doi.org/10.1177/0093650216644026>
- Hameleers, M., & Van der Meer, G. L. A. (2019). Misinformation and polarization in a high-choice media environment: How effective are political fact-checkers? *Communication Research*. <https://doi.org/10.1177/0093650218819671>
- Humphrecht, E. (2018). Where 'fake news' flourishes: A comparison across four Western democracies. *Information, Communication & Society*, <https://doi.org/10.1080/1369118X.2018.1474241>
- Jeong, S.-H., Cho, H., & Hwang, Y. (2012). Media literacy interventions: A meta-analytic review. *Journal of Communication*, 62(3), 454–472. <https://doi.org/10.1111/j.1460-2466.2012.01643.x>

- Jones-Jang, S. M., Mortensen, T., & Liu, J. (2019). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*. doi:0002764219869406
- Kahne, J., Lee, N.-J., & Feezell, J. T. (2012). Digital media literacy education and online civic and political participation. *International Journal of Communication*, 6, 1–24. <https://doi.org/1932-8036/201200050001>
- Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2017). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research*, doi:0093650217719596
- Lefevere, J., De Swert, K., & Walgrave, S. (2011). Effects of popular exemplars in television news. *Communication Research*, 39(1), 103–119. <https://doi.org/10.1177/0093650210387124>
- Marwick, A., & Lewis, R. (2017). Media manipulation and disinformation online (pp. 1–104). *Data and Society Research Institute*. Retrieved September 18, 2018, from <https://datasociety.net/output/media-manipulation-and-disinfo-online/>
- Nyhan, B., Porter, E., Reifler, J., & Wood, T. (2019). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability. *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>
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x>
- Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news”. *Digital Journalism*, 6(2), 137–153. <https://doi.org/10.1080/21670811.2017.1360143>
- 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>
- Tully, M., Vraga, E. K., & Bode, L. (2020). Designing and testing news literacy messages for social media. *Mass Communication and Society*, <https://doi.org/10.1080/15205436.2019.1604970>
- Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stępińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheaffer, T., & Stanyer, J. (2017). Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association*, 41(1), 3–27. <https://doi.org/10.1080/23808985.2017.1288551>
- Vraga, E. K., Bode, L., & Tully, M. (2020). Creating news literacy messages to enhance expert corrections of misinformation on Twitter. *Communication Research*, <https://doi.org/10.10093650219898094>
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3), 423–441. <https://doi.org/10.1080/03637751.2018.1467564>
- Wardle, C. (2017). Fake news. It's complicated. *First Draft*. Retrieved January 22, 2019 from <https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79>
- Wood, T., & Porter, E. (2018). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*. <https://doi.org/10.1007/s11109-018-9443-y>.