# The state-of-the-art in combating mis- and disinformation

*Lessons from pre- and debunking approaches*

Hameleers, M.

**DOI**
10.4324/9781003387404-3

**Publication date**
2024

**Document Version**
Final published version

**Published in**
Disinformation Debunked

# 1 The state-of-the-art in combating mis- and disinformation

## Lessons from pre- and debunking approaches

*Michael Hameleers*

### 1.1 Introduction

Mis- and disinformation may have severe ramifications for society and democracy at large. Whereas misinformation refers to unintentional untruthfulness (e.g., Wardle & Derakhshan, 2017), disinformation can be understood as the goal-directed creation and spread of intentionally deceptive information (e.g., Bennett & Livingston, 2018). As false information – irrespective of its intentions – may have harmful consequences, such as the amplification of distrust and polarisation, it is important to understand how mis- and disinformation can be combatted or prevented. Against this backdrop, this chapter will review the state-of-the art in literature on corrective information, with a specific focus on how pre- and debunking approaches have been found to be effective under different conditions. Based on this review, the chapter will end with a forward-looking perspective on how these different tools can be used effectively to both prevent and fight the uncontrolled dissemination of mis- and disinformation across the globe.

Pre-bunking approaches that prevent misperceptions and instil resilience among audiences are found to be effective (e.g., Roozenbeek & van der Linden, 2019; Tully et al., 2020). Here, pre-bunking is understood as any intervention to prevent or combat mis- or disinformation *before* it is disseminated, with the aim to make audiences more resilient and better equipped to resist the persuasive impact of false information. In this regard, inoculation strategies that expose people to a small 'dose' of false information before an actual threat are found to be effective in empowering citizens to distinguish false from true information (e.g., Roozenbeek & van der Linden, 2019). Likewise, media literacy interventions that offer citizens specific suggestions on how they can rely on critical skills to recognise misleading and false information are found to have a negative impact on the credibility and acceptance of mis- and disinformation (e.g., Tully et al., 2020).

Research on debunking initiatives, mostly understood as fact-checks that respond to mis- and disinformation *after* its spread, point to similar optimistic conclusions: Although important sources of resistance may be identified (Hameleers & van der Meer, 2020), fact-checks that offer a short, argument-based

DOI: 10.4324/9781003387404-3

refutation of falsehoods are found to be successful in correcting false information (e.g., Nyhan et al., 2019). Despite the indications on the effectiveness of pre- and debunking initiatives in extant academic research, we know little about the real-world impact of these interventions, or their combined influence in combating false information in current digital and fragmented information environments. Perhaps even more important, most literature has failed to consider the longer-term and unintended side effects that interventions may have for the acceptance and trust in real, authentic information.

Against this backdrop, this chapter not only offers an overview of the evidence that pre- and debunking interventions may work but also considers how they can be implemented and integrated in complex information ecologies, and how they can overcome the resistance and impediments of the technological and societal context in which they operate. Finally, this chapter considers how pre- and debunking should also consider the consolidation or recovery of trust in factually correct information that may easily be dismissed as mis- or disinformation in a post-factual information era.

## 1.2 The effects of mis- and disinformation

Disinformation is created with the intention to deceive, and make a predefined impact on people's beliefs (e.g., Hameleers, 2023). Misinformation may, in addition, unintentionally result in misperceptions among recipients (Wardle & Derakhshan, 2017). Considering that exposure to false information has been found to result in misperceptions, it is important to explore how these negative consequences can be prevented or remedied. But why are mis- and disinformation potentially harmful and effective forms of communication?

Extant research has shown that exposure to mis- and disinformation may, among other things, influence political perceptions (e.g., Hameleers et al., 2020), or people's trust in established information and the news media (e.g., Zimmermann & Kohring, 2020). Here, it is also important to note that experimental research shows that disinformation can be perceived as highly credible (e.g., Schaewitz et al., 2020). One explanation for the credibility of mis- and disinformation comes from the truth-default-theory (TDT) of information processing (Levine, 2014). This theory holds that, on average, people are more likely to perceive new information as honest than dishonest. Therefore, unless suspicion is actively triggered by information sources, content or style, mis- and disinformation are likely to be regarded as trustworthy (Hameleers, 2022). This truth bias can be exploited by malicious actors, who imitate credible news formats and authentic visual information as closely as possible to avoid triggering suspicion. This is especially the case with visual disinformation formats that are often used in online settings and can add persuasive power to deceptive narratives by offering 'proof' and a direct index of reality to false statements (Weikmann & Lecheler, 2022). Moreover, considering that mis- and disinformation are often repeated, they may come across as familiar and truthful information sources that can sidestep suspicion.

The often-delegitimising narrative of disinformation may undermine trust in real information and established media by foregrounding accusations of 'fake news' and misinformation, contrasted to alternative sources of allegedly trustworthy information (Hameleers, 2022). At the same time, discussions about the threats of mis- and disinformation and delegitimising accusations of false information may undermine trust in real information by disproportionally triggering suspicion (Van der Meer et al., 2023). For these reasons, it is crucial that pre- and debunking approaches to mis- and disinformation somehow maintain a balance between critical perspectives towards suspicious statements, without motivating a cynical view to all information encountered.

### 1.3 State-of-the-art in debunking techniques

Mis- and disinformation have been associated with severe ramifications for democracy, such as the delegitimisation of established knowledge, mounting relativism toward facts, polarisation and mistrust (e.g., Van Aelst et al., 2017; Waisbord, 2018). In this setting, the question on what should be done to correct and prevent misinformation has become more prominent than ever. From a journalistic perspective, fact-checking has been coined as an important response to the uncontrolled dissemination of misinformation. Here, we refer to fact-checks as stand-alone or embedded journalistic products and routines that offer verdicts about the level of (un)truthfulness of (political) information by checking dubious or prominent claims (e.g., Amazeen et al., 2015; Uscinski & Butler, 2013). Fact-checks can, first of all, be delivered by independent organisations, such as fullfact.org in the UK, Agence France Press in France, and factcheck.org in the US. In addition, fact-checks may be part of established journalistic formats, for example, in the form of a dedicated editorial section (i.e., *The Washington Post*'s Fact Checker section).

Irrespective of the format, fact-checking platforms are regarded as politically unbiased and neutral (e.g., Amazeen et al., 2015). Although these platforms may show a selection bias in their inclusion criteria, for example, by selecting only the most visible dubious claims, they may refrain from attaching political judgements to their verdicts. That being said, in the bipartisan setting of the US, fact-checkers have been accused of demonstrating a liberal bias (e.g., Shin & Thorson, 2017). Research has further demonstrated that Republicans are more likely to reject and counter-argue fact-checks (Shin & Thorson, 2017), and that people are generally most likely to select fact-checks when they attack misinformation that is incongruent with prior beliefs (Hameleers & van der Meer, 2020). Thus, although they claim to be neutral and objective observers of the information ecology, fact-checks may be subject to mistrust and politically coloured interpretations. But why would fact-checkers be effective, and for whom are their effects (least) pronounced?

Basically, fact-checks should work as they present news users with a univocal verdict on the truthfulness of claims – using a factual and rational

argumentation presented in easy-to-comprehend formats (Lewandowsky et al., 2012). The working of corrective information can also be understood through the perspective of the Truth-Default-Theory (TDT) coined by Levine (2014). The TDT states that humans are, as a default coping mechanism, wired to accept the honesty of the information they receive from others. As most information does not contain lies or misinformation (also see Acerbi et al., 2022), this default state of information acceptance should be valid most of the time, and helps news users to effortlessly navigate their information setting. Against this background, fact-checks should be effective in correcting misperceptions as they signal deception, and flag suspicious content to recipients. In other words, exposure to fact-checking information should prime suspicion, which should result in a deviation from the truth-default state.

There is ample evidence that exposure to fact-checking information lowers misperceptions, resulting in the successful debunking of misinformation. Meta-analyses, for example, show an overall positive and substantial effect of exposure to fact-checking information on the reduction of false beliefs (Chan et al., 2017; Walter et al., 2020). Even though some studies have pointed to negative or even 'backfire' effects of fact-checks (e.g., Nyhan & Reifler, 2010; Thorson, 2016), more recent experimental studies point out that exposure to fact-checking information lowers misperceptions (e.g., Hameleers & van der Meer, 2020; Nyhan et al., 2019).

Despite the potential of fact-checks, there are different limitations to their effectiveness. First of all, fact-checking information is most effective among recipients whose beliefs do not align with mis- or disinformation (e.g., Walter et al., 2020). In addition, fact-checking information is most likely to be selected by people who are already opposed to the claims of the debunked misinformation (Hameleers & van der Meer, 2020). Related to this, in an information era where facts are often disputed and mainstream media and journalism distrusted, fact-checkers may be regarded as biased, untrustworthy or even 'fake news' peddlers by growing parts of the population that have tuned out of established information sources (Primig, 2022). Here, it should also be noted that most research concluding that fact-checks are effective are based on artificial experimental study designs – which begs the question about fact-checks' effectiveness in real life media settings.

All in all, we can conclude that fact-checking information faces different challenges. Although people strongly inclined to believe mis- and disinformation may need fact-checking information most, they may be less likely to expose themselves to it and take the verdict into account. Even more so, their alternative media diets and distrust in established information sources may mean that they do not even come across debunking information. For this reason, it is important to present fact-checking information in a format that can overcome resistance and enhance the likelihood of selection among distrusting segments of the audience. But what kind of effective formats have been suggested and tested in extant literature?

First of all, the use of explicit rating scales and verdicts may enhance the clarity of fact-checks. Importantly, fact-checks should not emphasise and repeat the false statements that are debunked, as repetition of false information can enhance its credibility (Pennycook et al., 2018). A suggested format of fact-checking is therefore to present the verdict and arguments for the rejection of misinformation in a clear and accessible manner (e.g., Graves, 2017). It is also relevant to present the verdict early in the correction. Hence, in an overburdened information ecology, news users spend only very little time on reading news articles and background stories. Having only limited processing resources at their disposal, they may look for easy-to-process cues or heuristics. Responding to this bias, fact-checks may be most effective when they present recipients with a concrete and repeated verdict early on in the message, potentially supplemented with a visual rating scale that illustrates the level of untruthfulness (see also Amazeen et al., 2015, on the effectiveness of visual rating scales). A concrete example of such clear rating scales is the use of the truth-o-meter used by PolitFact.[1]

Second, as fact-checking information may be avoided by news users who need them most – such as citizens with stronger anti-establishment or disinformation perceptions – extant literature has experimented with different formats of fact-checking that should overcome resistance and selective avoidance. Hence, as literature shows that many citizens perceive the verdict of fact-checkers as biased or hostile against their views (e.g., Brandtzaeg & Følstad, 2017; Shin & Thorson, 2017), it is important that people do not perceive the correction as an attack on their beliefs. One way to overcome resistance and hostile perceptions towards fact-checks is to use humour or satire. As these formats are less likely to be associated with persuasive intent and political agendas, they may overcome the belief of political bias and hostility. In line with this, evidence shows that satirical refutations can be more effective than regular fact-checks among segments of the audience that are aligned with misinformation's false statements (Boukes & Hameleers, 2020). In a similar vein, Young et al.'s (2018) findings show that using humour in videos can enhance the effectiveness of corrective information.

Third, extant literature has offered indications that the source matters. More specifically, using regular source cues to present fact-checking information, such as established news sources and independent fact-checking platforms, may cause resistance among people who distrust such formats. As these citizens are most likely to encounter mis- and disinformation whilst avoiding regular established journalistic formats, it is important to find ways to reach them with corrections. One strategy proven to be successful is to use fictional cartoon characters or ordinary citizens to present and promote corrective information (Opgenhaffen, 2021). Fact-checks presented by cartoon figures did not cause resistance, and were overall perceived as credible and accurate sources of corrections. In a later application, the cartoon figures were replaced for a realistic fictional avatar called 'Anna'. Although more research is needed to explore the effectiveness of using alternative sources on

social media in direct response to false information, it is plausible that presenting corrective information on the same channels as misinformation and via non-elitist sources that resemble regular news users may overcome an important source of resistance and avoidance.

Taken together, extant research has shown that fact-checks may offer an influential remedy to mis- and disinformation in the digital age. Although there are important pitfalls and barriers that may limit its scope and effectiveness, there is convincing evidence that fact-checks can enhance resistance to mis- and disinformation, albeit not all segments of the population may be reached. Considering that false information spreads at a way faster rate than corrections, which are also extremely costly and time intensive to make, it is important to look beyond debunking approaches. An important question is therefore how resistance to mis- and disinformation can be enhanced among news users *before* they are confronted with false information in their newsfeed.

### 1.4 State-of-the-art in pre-bunking techniques

The limited reach of fact-checks, at times, and the potential barriers to their selection may be partially remedied by relying on a more elaborate toolkit of fighting misinformation. One major alternative or supplement to fact-checking is pre-bunking. As an umbrella term for interventions that aim to instil resilience to the effects of misinformation, 'pre-bunking' involves all techniques and approaches to prevent the negative impact of misinformation before people are exposed to misleading content. The envisioned effectiveness of pre-bunking can, just like the effect of fact-checks, be understood as a consequence of priming suspicion and herewith motivating a deviation from the truth-default state (Levine, 2014). In other words, pre-bunking makes news users aware of deception in their information environment, which should motivate them to critically assess the veracity and biases of information.

Pre-bunking has mostly been studied in the form of media literacy interventions (e.g., Hameleers, 2020; Vraga et al., 2020) or 'inoculation' techniques (e.g., Roozenbeek & van der Linden, 2019). Media literacy interventions typically contain a stepwise description about how people can detect misinformation, offering concrete recommendations to news users navigating their information environment (e.g., Vraga et al., 2020). Inoculation also offers a warning about false information and exposes people to a small 'dose' of misinformation that should protect them from the actual harms of misinformation (Roozenbeek & van der Linden, 2019). Hence, encounters with smaller doses of the threat helps people to recognise misinformation and enables them to defend themselves when actual threats of deception arise.

Recent experimental research has shown that pre-bunking techniques can be effective in enhancing resilience towards misinformation. Roozenbeek and van der Linden (2019) show that inoculation with 'fake news games' – in which people are confronted with misinformation as producers and consumers – helps news users to accurately identify misinformation presented to

them later. In a similar vein, experimental studies in which people are exposed to a media literacy intervention that tells them how to detect fake news lowers the credibility and believability of misinformation (e.g., Guess et al., 2020; Hameleers, 2020; Vraga et al., 2020). In a similar vein, using a pre-bunking intervention in education, Nygren et al. (2021) show that pupils were better equipped to detect false visual information after a two-hour intervention, whereas the credibility of real news was not affected. Although experimental research comes with clear drawbacks – such as the artificial and forced exposure setting and the short time frame between treatment and measurements – there is ample evidence that exposure to pre-bunking messages or games can help people to detect misinformation.

However, the priming of suspicion that is part of pre-bunking messages comes with an important drawback. Hence, by making people suspicious towards 'fake' news and deceptive content, a disproportionately cynical outlook on media content can be cultivated. Although misinformation is estimated to play a very small role in people's newsfeeds in terms of prominence (Acerbi et al., 2022), the focus on fake and deception in media literacy interventions may also lead news users to identify authentic and factually accurate information as misinformation. In support of this, Hameleers (2020) found that, although media literacy interventions lowered the credibility of misinformation, they also had a negative spillover effect on the credibility of authentic news. Modirrousta-Galian et al. (2022) also found that gamified inoculation treatments lowered the credibility of real information, and therefore did not improve the discernment between factually accurate information and misinformation.

These findings emphasise the importance of striking a balance between stimulating critical media literacy skills and maintaining trust in factually accurate information. Thus, although pre-bunking needs to help citizens to detect misinformation, such interventions should not overemphasise the threats of misinformation. Even more so, trust in factually accurate information needs to be consolidated, which means that the discernment between factually accurate information and misinformation is the targeted outcome instead of the detection of misinformation (Guess et al., 2020). This means that, looking beyond the effects on the recognition of just false information, interventions should also take into account the effects of interventions on trust in real news or factual information (also see e.g., Nygren et al., 2021). Against this backdrop, it is meaningful to approach the intended consequences of interventions as the extent to which news users are better equipped to make critical and well-informed decisions on the veracity and honesty of the information that surrounds them in today's overburdened information ecologies.

### 1.5 Suggestions for effective applications of remedies

At this stage, it can be concluded that both fact-checking information and pre-bunking techniques are proven to be effective remedies against misinformation. Yet, both routes towards more resilient societies come with

important drawbacks. Does this mean that we need to abandon these interventions altogether? The answer is negative. The identified barriers to interventions' effectiveness should be regarded as challenges that future research and applications need to combat and overcome. It shows that we may need to revise existing interventions, carefully considering sources of resistance and negative side effects. This section more specifically forwards four concrete suggestions on how interventions may be improved in order to overcome the identified challenges. The main assumption here is that although no solution on its own is perfect or best able to combat misinformation, the combination of different tools and their strategic targeting to vulnerable audiences may make a crucial difference in society's resilience to false and deceptive information.

### *1.5.1 Use channels, sources and formats to overcome avoidance and resistance*

Both pre- and debunking messages may be perceived as biased and unreliable by segments of society that have lost faith in the established political order and the media (Primig, 2022). Ironically, distrusting segments of society may be attracted most to alternative and hyper-partisan media that are most likely to disseminate counterfactual narratives or disinformation. In addition, experimental research shows that fact-checks are most likely to be avoided by people whose perspectives align with mis- or disinformation (Hameleers & van der Meer, 2020). The problem we face is therefore that interventions are most likely to be rejected or avoided by segments of society that are most susceptible to misinformation.

Responding to this challenge, the first major suggestion is to present corrective information or media literacy interventions on less established platforms and channels. More specifically, interventions may be most effective if they are presented on the same platform as misinformation, and part of the same networks as citizens with mis- and disinformation diets. A concrete example would be to present fact-checks directly as false flags on the social media platforms that are more likely to contain misinformation to and target media literacy warnings to community pages, platforms and environments that are likely to be visited by segments of society that tend to be exposed to misinformation. Thus, to increase the likelihood of selection among distrusting citizens, it is meaningful to not only present fact-checks or media literacy messages on independent stand-alone platforms of established media but also embed warnings and corrections on platforms that contain false information.

It also matters how the correction is presented, and which sources forward the verdict and correction. Importantly, people may feel attacked if an elitist institution – such as a fact-checking platform associated with the mainstream media – claims that they are wrong. For this reason, corrective information may be presented as subtle warning messages coming from ordinary people or sources people may identify with (see the example of cartoon figures or ordinary citizens on social media by Opgenhaffen, 2021). Importantly, these

sources should avoid a direct attack on people's beliefs by stating that their perceptions on the world are wrong, false or even 'stupid'. A more subtle approach that points people to factually accurate information contrasted to false beliefs may be more constructive – overcoming the resistance caused by a direct attack or a label that states that people are wrong. Formats such as a truth sandwich – which focuses on factually accurate information instead of repeating false claims – may also be used to reach the same goal. Essentially, corrective information should be presented carefully and should avoid the impression that people feel that their identities are under attack by labelling recipients as misinformed, wrong or even ignorant.

Related to this, corrective information may acknowledge the drivers of believing misinformation or disinformation. Too often, corrections fail to respond to the reasons why misinformation may be believable. A concrete suggestion is to acknowledge contexts of insecurity, uncertainty or the emotions associated with certain developments or events. As an example, the COVID-19 pandemic was surrounded by a context of uncertainty, high stakes, fear and disagreeing expert voices. Corrective information, however, typically failed to acknowledge this context and the drivers of believing misinformation – herewith potentially neglecting disenchanted voices in society. Based on this, corrective information may be most effective when it acknowledges and relativises the underlying drivers of supporting and believing misinformed claims. Acknowledging that COVID-19 is surrounded by uncertainty and fears about the causes and treatment of the virus may help to engage people who typically avoid corrective information as they may attack their beliefs and identities whilst failing to resonate with their worldviews.

#### *1.5.2 Consolidate trust and motivate critical (but not cynical) media skills*

Considering that interventions may backfire by priming disproportionate levels of cynicism and distrust in factually accurate information (e.g., Hameleers, 2022), it is crucial to consolidate trust in factually accurate information. Therefore, instead of exclusively warning about the threats of deception, media literacy interventions and inoculation strategies need to relativise and contextualise the threats of mis- and disinformation embedded in an information ecology of mostly accurate information (Acerbi et al., 2022). In other words, current interventions that only offer tips on how to recognise fake news may cultivate disproportionate negative views on the information ecology and prime suspicion beyond a constructive level that helps them to discern between factually accurate information and misinformation.

Concretely, media literacy interventions and fact-checkers may offer suggestions on where recipients can find factually accurate and reliable information on different issues. A reading list of selected sources for different issues may help, such as the official sources of informing citizens about COVID-19. Moreover, interventions can relativise the threats of mis- and disinformation by emphasising that most information that people encounter in their

newsfeed is reliable. However, some claims may be false or deceptive – which calls for a critical (but not cynical) approach to information. In any case, interventions can offer details on where misinformation is most and least likely to occur – which relativises the negatively biased worldview that misinformation is all around.

Another way to avoid cynicism is to focus less on the falsehoods and deception and more on the actual factual reality in the verdicts of a refutation. Fact-checkers may, for example, not simply repeat false information and stress the deceptive nature in their rating, but rather focus on the facts and the actual situation or evidence for reality. In addition, instead of warning about the deceptive nature of information, inoculation and media literacy approaches may stress the need to be critical, motivating users to verify information themselves by consulting multiple sources and reflecting on their own biases. Essentially, resilience towards misinformation may be achieved by stimulating critical skills and the confidence that people themselves are able to navigate their information environment. This entails the skills needed to find the majority of factually accurate information whilst being sensitised to the deception that may occur at times.

### *1.5.3 Experiment with modalities, formats and tailoring*

Although many fact-checking platforms aim to present their verdicts in accessible and clear manners, it takes considerate amounts of attention and cognitive resources to process corrective information. Fact-checking platforms typically aim to be rigorous and precise, which may pose a challenge on the scarce resources and attention of recipients. The often-technical and textual elaborations of fact-checkers may exclude vulnerable segments of the audience that do not want to – or are unable to – invest reading time and processing resources in consuming corrective information. But what can we do to grab the scarce resource of attention in a competitive media environment where misinformation is often more engaging and easy to consume than corrections?

Fact-checkers may use visually attractive formats to present their verdict in an easy to comprehend manner that requires only limited reading time. A concrete example is to embed two-sentence verdicts on social media to flag suspicious content, which also responds to the first recommendation to circumvent avoidance and rejection from citizens most likely to be exposed to misinformation. The two-sentence verdict may clarify why statements are regarded as suspicious, combined with a brief statement of factually accurate information that can be contrasted to the flagged suspicious claims. Although fact-checking platforms may intend to offer a complete and well-founded verdict, not all citizens will be able or motivated to read the entire analysis and the reasons why certain claims are disputed. Therefore, it is suggested to include a 'read more' option or a link to the full research and refutation when outlining and embedding a short and easy-to-comprehend verdict on social

media or other platforms where misinformation is likely to be encountered. This way, fact-checking may still be based on a complete and thorough analysis of reality, whilst also presenting recipients with an easy to digest recommendation.

Another suggestion is to present fact-checks in a more attractive and engaging format, for example by using short videos of a refutation or infographics that show recipients why claims are disputed after a quick look. Although such refutations require even more effort and financial resources, they may grab the audience's attention. Such videos or infographics should not be presented as stand-alone products on separate platforms or websites, but should ideally be embedded in the response sections of false, disputed or suspicious content online. Videos may also be suitable to explain to news users how they can be more critical news users. Hence, an audiovisual demonstration on how to recognise reliable information whilst being sensitive to deception may be comprehensible for citizens at different levels of cognitive skills and processing resources.

#### *1.5.4 An integrated approach towards resilient societies*

It is important to emphasise that although the different responses to misinformation may only have limited effects, they may have a meaningful impact on resilience when combined, repeated, targeted and adjusted to developments in the digital age. Although fact-checks may have a relatively limited reach, repeated exposure to debunking messages may also teach news users how to verify information themselves, and what types of narratives may be suspicious. The same applies to media literacy interventions and inoculation strategies. Not all messages may reach all citizens, but encountering warning flags and learning about the reasons why certain statements were retracted, dubious or unreliable may contribute to a more critical audience that is resilient to the threats of mis- and disinformation.

Moreover, when fact-checkers and media literacy initiatives offer clear guidance on where factually accurate information can be found, and how citizens can determine when information is trustworthy and reliable, we can avoid the negative side effects of strengthening distrust and cynicism related to trustworthy information. Overall, interventions need to offer the much-needed guidance for citizens to navigate complex digital information flows. Although most information is accurate and reliable, the digital setting affords manipulation and targeted attacks on people's beliefs. In addition, the many-to-many communication flows, the lack of gatekeeping and the low thresholds for access in the public domain also implies that people may be exposed to unintentional misinformation, or unverifiable claims that remain unchecked and unflagged.

This complexity of the information ecology needs to be central in education – not just in the educational message of pre-bunkers but also in educational programmes in schools and targeted digital literacy campaigns intended

for vulnerable groups (i.e., the elderly who may be reached through libraries or other institutions). Hence, the complex information flows require an understanding of misinformation embedded in the current information ecology, and approaches that only focus on the statement level or isolated suspicious narratives fall short in emphasising how the context of (deceptive) content matters for its identification and resistance to its influence.

## 1.6 What's next? A forward-looking perspective on corrective information

Although fact-checking and pre-bunking have been prominent strategies for multiple decades, new developments in digital communication and artificial intelligence may pose new challenges on the way forward. As deep fakes and other synthetic deceptive media are getting more widespread, corrective information may need to do more than just checking the accuracy and honesty of statements. Hence, they may increasingly need to verify the authenticity of materials that are seemingly real – moving beyond the correction of political statements and positions. Open-Source intelligence (OSINT) may offer an important aid in the process of tracing back the origins of synthetic materials, but the detection and creation of highly realistic deceptive media may always be entranced in an arms race. In this setting, it may become more important to teach recipients on how to detect discrepancies between reality and the viewpoints of known actors vis-à-vis the mediatised images of these persons. Simply pointing out technological glitches may be insufficient when technologies are improving rapidly. Therefore, the audience needs to be trained to critically scrutinise whether the depictions they observe are realistic. In this setting, it is helpful that news users are offered a toolkit to fact-check themselves, for example, by checking whether visual information is authentic (i.e., OSINT, Tineye, or InVID-WeVerify).

As misinformation is a relatively rare phenomenon compared to factually accurate information, future interventions may also need to focus on enhancing the credibility and trust in factually accurate information. As indicated by Acerbi et al. (2022), it may be more effective to enhance the trustworthiness of factually accurate information by 1 percent than to decrease the acceptance of misinformation to zero. Thus, the combined strategy of combating misinformation needs to be supplemented by strategies that increase the accessibility and comprehensibility of high-quality information. A main challenge in that regard is to lower society's resistance to established information sources and to motivate audience segments that avoid the news and mainstream media altogether. To enhance the appeal of high-quality trustworthy information, it may be needed to experiment with formats that are selected and accepted by disenchanted audience segments.

As part of this future endeavour, mainstream media, as well as corrective information initiatives, need to acknowledge and relativise the feeding ground of misinformation instead of rejecting or sidestepping the feelings and concerns of people likely to select or believe misinformation. Too often,

established media and other official sources of information ignore or even ridicule the concerns and feelings of disenchanted citizens – which may fuel polarisation and distrust to a more extreme level (Hameleers, 2022). Hence, considering the existing barriers for the selection of fact-checks and media literacy interventions among citizens that need them most, more resources have to be invested in building bridges and demonstrating higher levels of mutual understanding. Finding common ground between fragmented and opposed segments of society (i.e., everyone is concerned about the consequences of COVID-19/there is a lot of uncertainty and unclarity in official information) may contribute to enhancing trust in official information among citizens that are likely to avoid established information sources.

Related to this, more transparency on how knowledge is created and subject to high levels of uncertainty, change is needed to involve citizens in the process of knowledge created. For most citizens, the process underlying the formation of knowledge remains a black box, which also means that sudden changes in perspectives (i.e., masks were first deemed unnecessary and later obliged during the COVID-19 pandemic in many countries) are left unexplained. Only stating that decisions and guidelines suddenly changed without explaining why may fuel distrust in society and leave explanation room open for alternative narratives, conspiracies and misinformation. Explaining how decisions, policy and guidelines are changed and why factual interpretations may shift as expert knowledge is subject to change may leave less room for misinformation's alternative explanations of events.

In the way forward, transparency may also be more central to the platform's policies and guidelines. At this stage, there is only little knowledge about how algorithms curate and amplify misinformation and how financial interests of commercial platforms may promote the spread of falsehoods. In addition, there are currently no clear formal guidelines in place for the curation of false information. We are therefore in need of a stronger legal framework that explicates the guidelines for content producers and disseminators. Related, content policies and digital services acts need to work towards a more transparent digital society, where users know why they see certain content, and how their media diets are subject to mechanisms that are fueled by (commercial) interests.

## 1.7 Conclusion

Misinformation has been regarded as a severe problem that may undermine deliberative democracy. Extant research has demonstrated that both fact-checking and media literacy interventions can offer viable treatments, although they come with drawbacks. This chapter aimed to foreground the conditions under which interventions may (not) work and which strategies and adjustments may enhance their effectiveness. First of all, this chapter concluded that although fact-checks and other interventions are effective in artificial lab settings (e.g., Walter et al., 2020), they face challenges of

selective avoidance and rejection in real life. Hence, as growing segments of society are losing trust in established information sources – including journalists and fact-checkers – corrective information may not reach the most vulnerable users that are likely to approach misinformation that reassures their distrust in established sources.

Different strategies may be employed to increase the likelihood of selection and acceptance of interventions. For example, fact-checks can use less elitist sources, and experiment with sources that are more relatable for recipients. For example, it may be useful to use ordinary citizen sources that warn fellow social media users about suspicious content (Opgenhaffen, 2021), which may overcome resistance to fact-checking information. In any case, it is important to circumvent the growing levels of cynicism related to established information sources when correcting misinformation. Humour, ordinary citizen cues and argumentation styles that do not attack people who are inclined to support misinformation are recommended formats to present corrective information. Importantly, as different segments of society may have different (media) preferences, ideological biases and identities, the affordances of digital infrastructures may be used to tailor and target fact-checking information to recipients' preferences.

Beyond these specific recommendations, we can conclude that there is no single perfect intervention or solution to combat misinformation. Different interventions, such as fact-checking, media literacy messages, inoculation games or legal interventions, may all come with their own strengths and weaknesses. Individually, their effects may be limited. However, the combination of a mixture of interventions that are repeated, tailored and amplified may offer the best response to misinformation (see Hameleers, 2020). The combined impact of a well-balanced strategy based on a set of interventions may be bigger than the sum of their parts. For example, media literacy interventions may teach people how to be critical, and how to rely on fact-checking information when in doubt. Fact-checks, in addition to the general warnings offered by media literacy interventions, can offer specific guidelines for how the trustworthiness of statements may be rated.

Related to this, we have to critically ask what the desired outcome of fact-checking and media literacy interventions are: When is an intervention deemed effective? More specifically, reducing misperceptions in the short term may not be the desired endpoint or impact of corrections. Rather, to truly enhance resilience towards misinformation, citizens themselves need to have the toolkit to judge the veracity of information, whilst maintaining trust in factually accurate information. In line with this, interventions used to respond to misinformation should instil critical literacy skills that help people to discern between the majority of factually accurate and trustworthy information, while being sensitive to dubious and suspicious claims that warrant further attention. Ideally, critical news users are able to become fact-checkers themselves, and the interventions used to combat misinformation should offer them the tools they need to recognise when information is

in need of verification. Here, maintaining and enhancing trust in reliable information may be even more important than lowering the acceptance of misinformation (Acerbi et al., 2022).

Despite these recommendations, we are in need of more research on the effectiveness of corrections in the real world. Most evidence-based recommendations are currently founded on lab experiments or artificial survey settings in which people voluntarily contribute. These studies may miss the most vulnerable audience segments, and citizens that are most likely to select misinformation and avoid fact-checking information specifically. Therefore, we are in need of field experiments, as well as qualitative and interpretive methods that engage with citizens that resist fact-checking information, and studies that use real-life monitoring of different interventions. Moving forward, the conclusions of such studies can be used to further redesign and improve the effectiveness of current interventions proven to be worthwhile.

Taken together, we can conclude this chapter by emphasising the importance of ongoing initiatives to fight misinformation. Although no solution is perfect on its own, all means that are proven to be effective need to be used as part of a global toolkit that aims to enhance society's resilience to the threats posed by misinformation – which includes its undermining impact on the trustworthiness of factually accurate information.

## Note

1 PolitFact.com

## References

Acerbi, A., Altay, S., & Mercier, H. (2022). Research note: Fighting misinformation or fighting for information? *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-87

Amazeen, M. A., Thorson, E., Muddiman, A., & Graves, L. (2015). *A comparison of correction formats: The effectiveness and effects of rating scale versus contextual corrections on misinformation*. American Press Institute. Retrieved April 27, 2022, from https://www.americanpressinstitute.org/wp-content/uploads/2015/04/The-Effectiveness-of-Rating-Scales.pdf

Bennett, L. W., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. https://doi.org/10.1177/0267323118760317

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

Brandtzaeg, P. B., & Følstad, A. (2017). Trust and distrust in online fact-checking services. *Communications of the ACM*, *60*(9), 65–71. https://doi.org/10.1145/3122803

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. *PsychologicalScience*,*28*(11),1531–1546.https://doi.org/10.1177/0956797617714579

Graves, L. (2017). Anatomy of a fact check: Objective practice and the contested epistemology of fact checking. *Communication, Culture & Critique*, *10*(3), 518–537. https://doi.org/10.1111/cccr.12163

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

Hameleers, M. (2020). Separating truth from lies: Comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication & Society*, 1–17. https://doi.org/10.1080/1369118X.2020.1764603

Hameleers, M. (2022). *Populist disinformation in fragmented information settings: Understanding the nature and persuasiveness of populist and post-factual communication*. Routledge.

Hameleers, M. (2023). Disinformation as a context-bound phenomenon: Toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination. *Communication Theory*, *33*(1), 1–10. https://doi.org/10.1093/ct/qtac021

Hameleers, M., Powell, T. E., Van Der Meer, T. G., & Bos, L. (2020). A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. *Political Communication*, *37*(2), 281–301. https://hdl.handle.net/11245.1/a27a4957-6fd9-4a8f-b6d2-e21185fddc43

Hameleers, M., & van der Meer, T. G. L. A. (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

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

Lewandowsky, S., Ecker, U. K., 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. http://www.jstor.org/stable/23484653

Modirrousta-Galian, A., Higham, P. A., & Seabrooke, T. (2022). Effects of inductive learning and gamification on news veracity discernment. *PsyArXiv*. https://doi.org/10.31234/osf.io/4wfds

Nygren, T., Guath, M., Axelsson, C. A. W., & Frau-Meigs, D. (2021). Combatting visual fake news with a professional fact-checking tool in education in France, Romania, Spain and Sweden. *Information*, *12*(5), 201. https://doi.org/:10.3390/info12050201

Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (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

Opgenhaffen, M. (2021). Fact-checking interventions on social media using cartoon figures: Lessons learned from "the Tooties". *Digital Journalism*, 1–24. https://doi.org/10.1080/21670811.2021.2011758

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Primig, F. (2022). The influence of media trust and normative role expectations on the credibility of fact checkers. *Journalism Practice*, 1–21. https://doi.org/10.1080/17512786.2022.2080102

Roozenbeek, J., & Van der Linden, S. (2019). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research*, *22*(5), 570–580. https://doi.org/10.1080/13669877.2018.1443491

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

Shin, J., & Thorson, K. (2017). Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *Journal of Communication*, *67*(2), 233–255. https://doi.org/10.1111/jcom.12284

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 & Society*, *22*(1), 22–46. https://doi.org/10.1080/15205436.2019.1604970

Uscinski, J. E., & Butler, R. W. (2013). The epistemology of fact checking. *Critical Review*, *25*(2), 162–180. https://doi.org/10.1080/08913811.2013.843872

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., Sheafer, 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

Van der Meer, T. G. L. A., Hameleers, M., & Ohme, J. (2023). Can fighting misinformation have a negative spillover effect? How warnings for the threat of misinformation can decrease general news credibility. *Journalism Studies*, *2023*, 1–21.

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/10.1177/0093650219898094

Waisbord, S. (2018). Truth is what happens to news: On journalism, fake news, and post-truth. *Journalism Studies*, *19*(13), 1866–1878. https://doi.org/10.1080/1461670X.2018.1492881

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(3), 350–375. https://doi.org/10.1080/10584609.2019.1668894

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking, Council of Europe Report. http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf

Weikmann, T., & Lecheler, S. (2022). Visual disinformation in a digital age: A literature synthesis and research agenda. *New Media & Society*, https://doi.org/10.1177/14614448221141648

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

Zimmermann, F., & Kohring, M. (2020). Mistrust, disinforming news, and vote choice: A panel survey on the origins and consequences of believing disinformation in the 2017 German parliamentary election. *Political Communication*, *37*(2), 215–237. https://doi.org/10.1080/10584609.2019.1686095