Visual Mis- and Disinformation, Social Media, and Democracy

Journalism & Mass Communication Quarterly
2021, Vol. 98(3) 641–664

DOI: 10.1177/10776990211035395
http://journals.sagepub.com/home/jmq


Introduction

The spread of mis- and disinformation is an increasing concern for democratic societies around the globe (Lazer et al., 2019). In modern high-choice media environments, exposure to misinformation can be harmful and can have negative consequences for democratic governance as well as trust in news media and journalism more broadly (e.g., Bennett & Livingston, 2018; Chesney & Citron, 2018; Nisbet et al., 2021; Ognyanova et al., 2020; Vaccari & Chadwick, 2020). Some scholars have even argued that we are entering a post-truth era with an alternative epistemology and thus an alternative reality, in which, for instance, former president Obama was not born in the United States and global warming is simply a Chinese hoax (Lewandowsky et al., 2017) rather than largely undisputed scientific phenomenon (Cook et al., 2016).

These ongoing "debates" highlight the importance of several distinct but related constructs. The first is the concept of "fake news," which refers to false or misleading information (Lazer et al., 2019; see Tsfati et al., 2020). Egelhofer and Lecheler (2019) proposed that fake news can be conceptualized as a two-dimensional phenomenon differentiating (a) *fake news genre* or "the deliberate creation of pseudojournalistic disinformation" (p. 97) from (b) the *fake news label* used (e.g., by politicians like Donald Trump) to delegitimize news media. Second, there is a distinction between disinformation and misinformation, whereas disinformation is a subset of misinformation. Disinformation is spread intentionally by various actors who know that the information is false. In contrast, misinformation is spread by actors who mistakenly believe the information to be factually correct when it is not (Vaccari & Chadwick, 2020; Vraga & Bode, 2020).

Previous research has primarily focused on textual forms of misinformation, while visual and multimodal forms (e.g., news images, memes, and videos) of misinformation have received much less attention. This is surprising because visual information may affect how media consumers select and process information (Zillmann et al., 2001; see also Garcia & Stark, 1991; Sargent, 2007). Furthermore, visual information can affect news consumers' emotional reactions (e.g., Iyer et al., 2014), attitudes (Matthes et al., 2021; Powell et al., 2015; von Sikorski, 2021; von Sikorski & Ludwig, 2018), and behavioral responses (Powell et al., 2015) independently of textual communication (von Sikorski & Knoll, 2019). This tendency is likely due to visual information coming "with an implicit guarantee of being closer to the truth than other forms of communication" (Messaris & Abraham, 2001, p. 217). Thus, visual mis- and

disinformation may be particularly persuading (see Messaris, 1997) and could have damaging effects for democratic governance. Visual information can be manipulated or taken out-of-context and can be (mis)used as a credible type of "proof" (e.g., "deepfakes" of politicians). Emerging research has shown that new multimodal forms of misinformation are disseminated quickly and seamlessly via social media and can have considerable negative effects on political attitudes and decision-making (Hameleers et al., 2020; Vaccari & Chadwick, 2020).

For instance, focusing on the period leading up to the 2019 Indian national elections, Garimella and Eckles (2020) showed that 13% of all images shared on WhatsApp public groups in India qualified as visual misinformation (for visuals in COVID-19 misinformation, see Brennen et al., 2020). In a bottom-up approach, ordinary users can "produce" and spread mis- and disinformation via social media on their own by manipulating photographs or by using simple editing techniques to manipulate original video material (slowing down a sound-track, de-/re-contextualizing visual information, etc.). However, mis- and disinformation can also be spread top-down. For instance, political actors can disseminate mis- and disinformation to their followers via social media and thus quickly reach large audiences, bypassing mainstream media outlets, using both nonsophisticated forms of mis- and disinformation (e.g., out-of-context visual information) and sophisticated manipulation techniques like "deepfake" videos based on artificial intelligence and machine learning procedures (Vaccari & Chadwick, 2020; for an example, also see Christopher, 2020). Although, bottom-up and top-down dissemination processes can generally be differentiated, multimodal misinformation may further spread through social media networks in complex ways, enabling political actors to circulate and further disseminate misinformation created by ordinary citizens or political groups (for an example, see Harwell, 2019). Yet, social media platforms are not the only media sources that influence whether and how multimodal mis- and disinformation spreads (Allcott et al., 2019; Donovan, 2021; Guess, Nyhan, et al., 2020). Tsfati and colleagues (2020) emphasized the importance of mainstream media in the spread of mis- and disinformation, as citizens regularly learn about political disinformation campaigns via mainstream media coverage.

Visual political mis- and disinformation is still not well understood by the scholarly community as scientific research about this phenomenon is still in its infancy—leaving many questions unanswered. For instance, how can visual misinformation be effectively debunked (Hameleers et al., 2020; Young et al., 2018)? How can backfire effects and continued influence effects of misinformation be best prevented (e.g., Lewandowsky et al., 2020; Nyhan, 2021; Stubenvoll & Matthes, 2021)? Are there ways to inoculate individuals against (visual) misinformation (Basol et al., 2021; Compton et al., 2021)?

The aim of the invited forum is to find answers to some of these questions and to bring together leading researchers from the fields of political communication, visual communication, psychology, and data science to provide a comprehensive overview of the state of research, noting key challenges and identifying avenues for future research. The forum brings together expert scholars focusing on key domains of visual political mis- and disinformation. Viorela Dan (University of Munich) focuses on

different types of disinformation videos and challenges for journalism and democracy. Britt Paris (Rutgers University) and Joan Donovan (Harvard University) examine online platform functionality and the fight against audiovisual disinformation. Michael Hameleers (University of Amsterdam) points out the effects of multimodal disinformation, and how it can be potentially debunked and efficiently corrected. Based on inoculation theory, Jon Roozenbeek and Sander van der Linden (University of Cambridge) examine an innovative way of prebunking or "vaccinating citizens against visual disinformation" before media users are exposed to multimodal falsehoods. Finally, I will point out "next steps" and future avenues for research on multimodal misinformation. In all, these contributions offer important insights and clarify why we should continue to research and expand our knowledge of multimodal mis- and disinformation in the future.

Christian von Sikorski
Assistant Professor of Political Psychology
Department of Psychology, University of Koblenz-Landau

Fake Videos: Challenges for Journalism and Democracy Emanating From Deepfakes and Cheapfakes

There is video "evidence" of Putin being on trial for corruption (Harding, 2012), Obama calling Trump "a total and complete dipshit" (BuzzFeed, 2018), and Queen Elizabeth II complaining about her family (Sawer, 2020). However, none of this "evidence" is real; in fact, it was entirely fabricated. The videos of Obama and the Queen used artificial intelligence (AI) and are what have become known as *deepfakes*. That of Putin used traditional video editing techniques and is what experts now call a *cheapfake* (Paris & Donovan, 2020). Whether deep, or cheap, all these videos are fake. Fake videos are audiovisual forgeries created purposely to suggest that someone did or said something that never occurred (see also Chesney & Citron, 2018; Nelson & Lewis, 2019). Such videos are "disinformation because they originate with intentional acts (the creation of the [fake] video). But they become misinformation, too, if circulated online by people who mistakenly believe them to be truthful representations" (Vaccari & Chadwick, 2020, p. 10).

Creating realistic-looking, yet fake, videos in which—for instance—politicians move like puppets on a string and utter words put into their mouths would have been unthinkable only a few years ago, but it is becoming easier each day (Stankiewicz, 2019, consider #refaceapp). To illustrate, with the help of a video agency, I was able to produce fake videos to use as stimuli in a study. Their high level of realism sent shivers down my spine, an impression corroborated by an AI expert in my network (who had asked to see one sample deepfake out of curiosity). Upon receipt, he asked me to double-check that I had sent him the correct version of the video. Obviously, he said, I must have mistakenly sent him the original (unaltered) file instead. I had not; he got a deepfake; it was just *that* good.

If researchers can create convincing fake videos so quickly and easily, wouldn't rogue actors, too? And, if they did, what would this mean for democracy, for journalism, and for shared knowledge among the public at large? Should these questions seem premature, consider the use of a deepfake in a local Indian election (Christopher, 2020) and the attempted coup in Gabon sparked by discussions over the credibility of an alleged deepfake spread by the regime to "demonstrate" that the president had recovered from a stroke when he had not (Cahlan, 2020). Consider also the statement of an expert in the detection of deepfakes: He reported that half a dozen politicians asked him to demonstrate their innocence by running incriminating videos through his algorithm (Breland, 2019). Perhaps, then, the apocalyptic undertones surrounding fake videos, deepfakes in particular, are appropriate (see Schwartz, 2018). After all, a plethora of research suggests that the average viewer is as unlikely to question something they see in a realistic-looking video as they are to doubt something they have seen with their own eyes (Chesney & Citron, 2018; Dan, 2018).

Obviously, determining the consequences of fake videos in the political domain is at its core an empirical question. From all the threats recounted in the literature, two broad categories seem particularly relevant (see Chesney & Citron, 2018; Vaccari & Chadwick, 2020). First, fake videos can ruin the reputation of the individual involved. For instance, in the case of a political candidate, a fake video could affect public attitudes and threaten the individual's political success. Second, detrimental spillover effects can arise, such as a generalized distrust in social and political actors and a sweeping sense of confusion over what is real and what is not. Such effects might be mediated by perceived realism and moderated by visual and digital literacy, among others. Moving forward, these are the key variables that scholars should focus on. But effects are certainly not the only thing we should devote attention to. Rather, we must also study the extent to which humans and algorithms succeed in recognizing fake videos, how and why these videos spread online, and how negative effects could be corrected. Existing evidence, while still preliminary, paints a bleak picture. It suggests that astute fake videos are more likely to spread fast and deep, and thus may be able to yield greater effects that are particularly difficult to correct (Ahmed, 2021; Dobber et al., 2021; Rössler et al., 2018; Vaccari & Chadwick, 2020).

But most of all, as soon as evidence on this detection—diffusion—effects triad begins to accumulate, we would be best advised to start thinking about building theory. After all, the standard research formula (in which we "identify a potential toxic media effect → document the effect with research findings → ameliorate the noxious effect") has not served us well in the past (Berger et al., 2010, p. 3). Rather than chasing one presumed threat after another, the scholarly community can instead help develop feasible solutions to this increasing social problem.

The road ahead for those seeking to address the threats posed by fake videos is likely to be long and full of obstacles. Although there is value in technical solutions—focused on (a) demonstrating whether something actually happened or not, such as lifelogging (Chesney & Citron, 2019), or (b) automatically identifying disinformation hoping to decontaminate the communication ecosystem—this may not be sufficient. A parallel approach might be to focus efforts on learning how to live with fake videos,

understanding the threats they pose, and increasing digital literacy in algorithmic spaces. Specific suggestions include concentrating on how certain actors can mitigate the effects of fake videos. For instance, journalists can contribute to this mitigation effort through gatekeeping and fact-checking, whereas digital literacy educators can help by prioritizing visual aspects. As communication scholars, we are predestined for research informing such efforts, and it seems wise to push fake videos to the top of our research agenda.

Viorela Dan
Postdoctoral Researcher (Akademische Rätin)
Department of Media and Communication, LMU Munich

Long on Profit and Years Behind: Platforms and the Fight Against Audiovisual Disinformation

Neural networks and generative adversarial networks have long been able to generate realistic videos that never happened, but prior to 2017 had been relegated to major motion picture studios for the enjoyment of mass audiences or in computer science research labs concerned with "computer vision." In 2017, consumer-grade, or sometimes free, image manipulation software using machine learning gained public attention, as porn videos appeared on Reddit with faces of famous women like Gal Gadot and Scarlett Johanssen grafted on porn actors' bodies (Cole, 2017). Since then, an app for creating "deep nudes" of anyone's picture was developed, made widely accessible, then almost immediately shuttered, when the developers suddenly understood the harm that could come from it. Mysteriously manipulated nude images from the app materialized on encrypted Telegram conversations over a year later. Although these examples do not feature political figures, pornified deepfakes of political figures like Alexandria Ocasio-Cortez and Nancy Pelosi abound online. But more to the point, audiovisual content is often a medium for spreading dominant political ideologies like white supremacy and misogyny without featuring political figures.

These examples of rather technically sophisticated "deepfakes" made by amateurs online draw attention to the increasing prevalence of image-based informational objects, generated for expression, for play, or for experimentation. Their creation and spread compels political questions around ethics and policy that require grappling with how structural power is reified through contemporary information and communication infrastructure. While popular discourse guised as critique suggests that the dangers posed by deepfakes simply require new technologies of information security and verification, it misses the point. The worry over deepfakes primarily lies in the realm of electoral politics and creating widespread panic surrounding certain groups or major political events and not in more mundane uses of audiovisual technology that have been historically wielded to silence people whose interests are never considered in platform policy (see Citron, 2016; Franks, 2018; Noble, 2018), whereas white-supremacist sites are allowed to spread harmful misinformation by being filtered to the top of search results. Moreover, the focus on technical solutions creates economic openings for technology companies to

deploy technical detection systems for deepfakes while shirking their responsibility for the harm caused by cheapfakes produced through simple editing techniques, or even basic text-based disinformation (Paris & Donovan, 2020).

We suggest that the proliferation of audiovisual fakes generated from sophisticated machine learning models that can be accessed by amateur communities online, or deepfakes, and "cheapfakes," that are produced through conventional methods of editing video with free software, requires more than technical solutions to address information vulnerabilities and negative social consequences. We suggest that we need to dismantle current communication and information infrastructures that profit from promoting dangerous content and creating new modes of pro-social infrastructure. Particularly, the entire data science as a field has never reckoned with the fact that most big data sets are nonconsensual and, in the case of deepfakes, exploit women at a much higher rate than men (Adjer et al., 2019). What would it take to reimagine consent in an era of AI that requires massive data stores to function?

Long before the advent of computer-generated images in film, recontextualizing images under the guise of evidence spread across via photographs, film, and video and it continues today in social media (Abel, 2004; Attwood, 2007; Coopersmith, 1998). The way audiovisual impersonation technologies work today make anyone with an online profile and a few images of themselves online, or even in their phone, fair game to be faked. But we know that there are specific groups of folks who are more in harm's way than others. There are many instances of manipulated images and videos are already wielded by amateur communities online to target women, LGBTQIA (lesbian, gay, bisexual, transgender, queer and/or questioning, intersex, and asexual and/or ally) folks, people of color, and those questioning powerful systems (Citron, 2016; Franks, 2017, 2018; Noble, 2013, 2018), online movements, harassment, and misinformation (Jones, 2019; see Stop Online Violence Against Women, 2018), and others studying hate and racism online (Daniels, 2009; McGlynn et al., 2017; Nakamura, 2002).

To address the hype around deepfakes, federal and state legislation has been introduced and passed in a few cases, especially related to image-based abuse in the narrow form of revenge porn (Clarke, 2019; Cyber Civil Rights Initiative, 2020; Sasse, 2018). These laws do little to address cheapfakes and textual misinformation writ large. Ironically, these laws punish users who generate and spread the fakes, but do not address platforms who make money from spreading this content (Paris & Donovan, 2020).

Platforms have instituted various policies to take down technically sophisticated deepfakes that interfere with the "political process" but not "works of art, parody, or satire" with little information about what constitutes any of these categories (Bickert, 2020). They have introduced labeling systems for text-based disinformation that do little to dismantle the problem of disinformation (Ognyanova, 2021; Ognyanova et al., 2020). Deepfakes and text-based disinformation are easier to recognize through technical methods because it is easy for a machine to read differences in pixels between the source video and the grafted face or body. This may or may not be why we have seen very few deepfakes of elected officials that trigger panic or doubt over political processes spread on these platforms with these policies.

All the while, cheapfakes around the political process proliferate to dangerous ends, especially recontextualized media that involve circulating clips out of context that intentionally distort the intended meaning of the speaker (Dreyfuss, 2020). Most famously, a slowed down video of Nancy Pelosi video was shared millions of times, appearing to show her drunk in public (Paris & Donovan, 2020). When it was revealed that this was an edited video, platform companies did not take serious action to remove it. These technical methods still cannot detect cheapfakes because their data structures are the same as any other video.

Text-based misinformation, too, routinely falls through the cracks because truth and falsehood are not discrete categories that can show matches with certain words or phrases with natural language processing algorithms, or pixel irregularities in images. With text-based disinformation, truth and falsehood are dependent on social practices of interpretation that often elude technical models.

The volume and speed of disinformation proliferating through online platforms are a source of enormous profits for those online platforms. Moving forward, until we dismantle and rethink the infrastructure used to produce and disseminate audiovisual fakes, which includes the social, political, and economic practices around technical systems, these problems will not change. Platforms hire humans to do the mind-numbing work of content moderation using an "algorithm" or a set of rules that allow for interpretation of social and political contexts after it has been flagged by a person, but these companies profiting from the spread of content pay content moderators poorly and treat them with flagrant disregard (Roberts, 2019).

It is the human aspect of these systems of verification that hold the greatest promise, but it will entail paying and training moderators properly for their work of upholding the reputation of truth, if not that of the platforms. Librarians and archivists are experts at vetting information and indeed have been building information systems for the public interest for a century. Moreover, libraries offer a unique opportunity and locus of further research on combatting disinformation as they are trusted, localized sites of information negotiation that can engage the public in-person discussions around interpretating and evaluating information (Geiger, 2017; Sullivan, 2019). There are federal agencies like the Federal Trade Commission and existing state and federal laws, like torts, that could dismantle the antisocial practices of platforms enacted because disinformation has not blossomed into a fully functioning industry. There exist a number of ways forward, but it is clear that we must mobilize political will to change these information and communication infrastructures before they become too big to legislate.

Britt Paris
Assistant Professor of Library and Information Science
School of Communication and Information, Rutgers University

Joan Donovan
Research Director
Harvard Kennedy School's Shorenstein Center on Media, Politics, and Public Policy

The Effects of Visual Disinformation and Debunking Falsehoods: State-of-the-Art and a Future Research Agenda

Techniques to alter, manipulate, or doctor images or even videos are getting more sophisticated, widespread, and accessible to nonprofessional communicators and politicians (e.g., Dobber et al., 2021; Paris & Donovan, 2020). As visuals may amplify framing effects and the perceived authenticity of information (Hameleers et al., 2020; Powell et al., 2015), visual communication plays a central role in the current digital disinformation order and may augment the effects of disinformation on message credibility, trust, issue agreement, or even political judgments.

Despite its importance, we know markedly little about the impact of multimodal disinformation (i.e., visuals and text, deepfakes, or cheapfakes) and the effectiveness of providing corrective information in a multimodal setting. We also currently lack a clear research agenda on how modality should be integrated in research on disinformation's impact. Therefore, this essay aims to (a) outline the state of the art in research on the effects of multimodal disinformation, (b) reflect on the effectivity of corrective information in response to multimodal disinformation, and (c) offer guidelines for future research that aims to disentangle the impact of multimodal disinformation on democracy.

The Scope of Multimodal Disinformation's Effects

Before delving deeper into the effects of multimodal disinformation, we need a clear working definition. I define multimodal disinformation as the practices involved in altering, de-contextualizing, doctoring, or fabricating (audio)visual materials with the intention to mislead receivers (also see Hameleers et al., 2020). This working definition recognizes the difference between mis- and disinformation, as pointed out earlier in this Invited Forum: Misinformation is false information that is not intentionally misleading (Vraga & Bode, 2020; Wardle, 2017), whereas disinformation is, by definition, about intentional untruthfulness (e.g., Bennett & Livingston, 2018; Freelon & Wells, 2020). Multimodal disinformation can combine text with visuals (i.e., memes, cropped images with a misleading caption) or may rely on audiovisual cues (i.e., cheap or deepfakes).

Disinformation may affect credibility, issue agreement, or even behavioral intentions due to what scholars call veracity bias in information processing (Lang, 2000). This bias implies that people are more likely to accept the veracity of incoming information than to deem it untruthful. This is especially relevant to consider in high-choice information settings. To navigate the overload of messages in digital information settings, citizens rely on heuristic cues instead of systematic processing of arguments—such as the formatting of information, sources, or presentation style. Disinformation may profit from this truth bias: False information is spread using the same formats as authentic news, and parts of reality are used strategically to tell lies—a technique that

is better known as paltering (Rogers et al., 2017). When disinformation is presented using similar formats as authentic information, inconspicuously embedded in hybrid media ecologies (Kim et al., 2018) and close to the truth (Stroud et al., 2017), manipulated information may be very convincing.

References to truthfulness and authenticity are central to multimodal disinformation effects (see Hameleers et al., 2020). Crucially, visuals are more attention-grabbing and emotionally engaging than textual information (Powell et al., 2015). For this reason, visual disinformation may have a stronger impact on political judgments than textual disinformation: Visuals help to transport news consumers in a storyline and can override systematic processing of (faulty) lines of argumentation (Hameleers et al., 2020). Another crucial quality of visuals is their more direct index of reality compared with text alone (Messaris & Abraham, 2001): Visuals bear a stronger relationship to the depicted reality than the abstract descriptions offered by text, and the richer and more vivid reality displayed in visual information should elicit stronger emotions and behavioral responses (Powell et al., 2015).

Based on this theoretical backdrop, recent empirical research has shown that multimodal disinformation is slightly more credible than textual disinformation (Hameleers et al., 2020). Applied to deepfakes, short (targeted) false videos were found to be highly credible and resulted in more negative evaluations of the depicted politician (Dobber et al., 2021). Vaccari and Chadwick (2020) show the impact of deepfakes on (dis)trust: Although their study shows that people are not likely to accept implausible statements as truthful, audiovisual manipulations cause confusion and distrust in online news. Overall, audiovisual disinformation can achieve some of the intended goals behind disinformation: to confuse the audience, offer credible counter-narratives, and foster distrust and cynicism in (legitimate) news. Despite these insights, we lack a clear understanding of how the impact of deepfakes can be compared with the effects of textual disinformation.

How to Correct Multimodal Disinformation

Because images tend to be perceived as more authentic, they may also be harder to correct. Yet, empirical research has not offered support for this assumption: Hameleers et al. (2020) found that fact-checkers are equally effective in correcting textual and multimodal disinformation. In addition, this study found that multimodal formats of fact-checking are not more effective than refutations based on text alone. However, Amazeen et al. (2018) found that visual rating scales may be more effective than text under certain conditions. In line with this, Nyhan and Reifler (2011) show that using graphic information in corrections is more convincing than corrections without a rating scale. Based on this, it seems that relying on a visual and easy to process rating of accuracy, alongside a systematic debunking of false statements, may contribute to the effectiveness of fact-checking. Future research needs to experiment more with different formats of incorporating multimodal information in fact-checking. An important unanswered question, for example, is whether a video message that systematically

debunks disinformation is more effective than the "false" flags and labels typically used by fact-checking organizations in online settings.

Where to Go From Here: An Agenda for Future Research

It is crucial to keep track of new developments in (audio)visual doctoring and incorporate these in future studies of media effects. Hence, to date, empirical research on disinformation's impact does not reflect the multimodal reality of online information environments. Against this backdrop, the following four recommendations for future research can be formulated:

1. *Take modality into account in research designs*. Experimental designs may and should explicitly incorporate modality as additional factor—varying the presentation style of false information (i.e., text alone, text plus image, video). Because it is extremely difficult to keep factors constant across different modalities (i.e., the background of a video or the voice of a depicted politician may bias results), extensive pilot testing is needed. In addition, researchers should share stimuli and techniques of multimodal doctoring to aid future research.
2. *Embed disinformation in realistic online environments*. Forced exposure experimental designs may overlook the crucial role of selective exposure, hybrid media ecologies, and algorithmic biases, all of which are crucial factors for disinformation's impact on society (e.g., Kim et al., 2018). These mechanisms may be taken into account in experimental research, for example, by offering participants the choice to select their preferred online story or by simulating the mechanisms of news recommenders and algorithmic biases in multiwave experiments.
3. *Focus on longer-term effects*. Although experimental research is useful, extant research has mostly mapped the impact of disinformation directly after exposure. Multiwave experiments that measure the delayed effects of disinformation and corrections may offer more realistic insights into the consequences of multimodal disinformation. In addition, pairing content analyses of multimodal disinformation with multiwave surveys and media exposure measures may offer insights that are more externally valid. Ideally, findings from experiments and panel studies will be combined to assess the long-term effects of multimodal disinformation.
4. *Make procedures and data available to fellow researchers*. Techniques to create deepfakes are developing rapidly, and improving at a fast pace. If we aim to directly respond to the real-life threats posed by these techniques—and develop evidence-based policy recommendations—we have to respond quickly. Therefore, it is crucial that researchers make scripts, techniques, and stimuli available to fellow researchers who study the impact of disinformation across disciplines, regions, and temporal variations (i.e., disinformation in routine vs. election periods, deepfakes on COVID-19).

Conclusion

Technological affordances play a central role in the construction and dissemination of disinformation. This essay focused on the role of multimodal manipulation in the effectiveness of disinformation and corrective information. (Audio)visual cues make disinformation more credible and can help to realistically embed false storylines in digital media ecologies. As techniques for (audio)visual manipulation and doctoring are getting more widespread and accessible to everyone, future research should take the modality of disinformation, its long-term effects, and its embedding in fragmented media ecologies into account.

Michael Hameleers
Assistant Professor Political Communication and Journalism
Amsterdam School of Communication Research (ASCoR)

Prebunking: Vaccinating Citizens Against Visual Disinformation

The study of misinformation has exploded in recent years with much focus on how fake news spreads on social media (Del Vicario et al., 2016; Johnson et al., 2020; Vosoughi et al., 2018), what determines people's susceptibility to fake news (Pennycook & Rand, 2019; Roozenbeek, Schneider, et al., 2020), and evaluations of interventions that might be effective in helping people to spot and resist misinformation (Fazio, 2020; Guess, Lerner, et al., 2020; Roozenbeek & van der Linden, 2019; van der Linden et al., 2021). Importantly, although the media increasingly relies on visuals, the problem of *visual* misinformation remains relatively understudied (Brennen et al., 2020; Hemsley & Snyder, 2018). Understanding how the use of visuals affects the perception and spread of misinformation is of key importance, as research has shown that people's perceptions of news stories are strongly influenced by what visuals are used (Zillmann et al., 1999) and that the acquisition of textual information is facilitated by imagery, especially emotional images (Zillmann et al., 2001). In fact, recent research on the "truthiness effect" highlights that a "non-probative photo can bias people to believe that an associate claim is true despite the fact that the photo offers no diagnostic evidence for the claim's veracity" (Zhang & Newman, 2020, p. 1).

The Power of Visual Disinformation

A concrete example of how visuals can aid the spread of misinformation is shown in Figure 1, which depicts a screenshot from a video posted on a Facebook group named "News World" in March of 2018 (van der Linden & Roozenbeek, 2020).

"News World" posted the video on its page alongside the claim that it showed Muslim immigrants attacking the Basilica of Saint Denis in Paris during mass. It gained traction quickly, amassing around 1.2 million views the day after it was posted, with politicians such as Front National leader Marine le Pen expressing outrage on social media (Le Pen, 2018). However, fact-checkers quickly pointed out a series of errors in the post (Damarla, 2018; Snopes, 2018). First, there was no evidence that the

Figure 1. "News World" Facebook post (March 20, 2018).
Source. Reprinted with permission from van der Linden and Roozenbeek (2020).

people in the video were either Muslims or immigrants to France. Second, members of the Saint Denis church clergy stated that it had not been "attacked," as the Facebook post claimed. Rather, it was the site of a demonstration against a proposed bill that would restrict immigrants' ability to obtain asylum in France. Third, the demonstration did not take place during mass. And finally, the police did not try to stop the protestors, but rather appear to have peacefully removed them from the premises about an hour after the start of the demonstration.

The above example shows the damaging potential of visually powerful disinformation, which is further complicated by the fact that it was not the video itself that contained false information; rather, it was the misleading *context* in the "News World" Facebook post that fuelled its virality. By the time the fact-checks went online, the damage was already done: The video had been watched by millions of people, many more than the fact-checks were likely to reach. Moreover, even if everyone who had originally been exposed to the video could also be shown the fact-check, misinformation is "sticky," meaning that corrections do not fully nullify belief in the original misinformation, a phenomenon known as the "continued influence effect" (Lewandowsky et al., 2012; Walter & Tukachinsky, 2020).

Prebunking: A Psychological "Vaccine" Against Misinformation

Given the challenges associated with fact-checking, our approach has shifted gears from traditional debunking to what we call "prebunking." Inoculation theory posits that exposing people to a weakened version of a persuasive argument creates mental "antibodies" against it, much like a medical vaccine triggers the creation of antibodies

against a pathogen (Compton, 2013; McGuire, 1964; McGuire & Papageorgis, 1961). In other words, by preemptively debunking (or "prebunking") misinformation, people are less likely to be swayed by it in the future. Meta-analyses support the efficacy of inoculation messages (Banas & Rains, 2010). Yet, to make inoculation theory scalable within the context of online misinformation, our research has moved away from the traditional issue-based approach to inoculation to focusing more on the techniques that underpin misinformation in general such as the use of moral-emotional language, impersonating people, and spreading conspiracy theories (Lewandowsky & van der Linden, 2021; Roozenbeek & van der Linden, 2019).

One of the ways to inoculate people against such techniques is through entertaining online games that visually simulate a social media environment and forewarn and expose people to weakened doses of these techniques in a controlled environment. We have developed three such games so far: *Bad News* (http://www.getbadnews.com), *Harmony Square* (www.harmonysquare.game, a game about political and electoral misinformation), and *Go Viral!* (http://www.goviralgame.com, a game about COVID-19 misinformation). All these games are choice-based and have multiple levels, each of which exposes a particular misinformation technique.

In a series of studies, we have demonstrated that playing an inoculation game reduces the perceived reliability of misinformation (Roozenbeek & van der Linden, 2019); increases people's confidence in their ability to spot misinformation (Basol et al., 2020; Basol, Roozenbeek, et al., 2021); and reduces their self-reported willingness to share misinformation with other people in their network (Roozenbeek & van der Linden, 2020). We have replicated these effects across different cultures (Basol, Roozenbeek, et al., 2021; Roozenbeek, van der Linden, & Nygren, 2020) and found that people remain significantly better at spotting misinformation for at least 2 months after playing when given regular "booster" sessions (Maertens et al., 2021).

Visual disinformation is explicitly addressed in our games, particularly in *Go Viral!*, as shown in Figure 2. *Go Viral!* simulates a person's gradual descent from a regular social media user to the moderator of a misinformation-spreading social media group called "Not Co-Fraid." In one of the levels, the player comes up with a conspiracy theory about a target of their choice (e.g., a government or an NGO). As their conspiracy gains popularity, players are given the option of adding fuel to the fire by sharing a video of an unrelated protest, making it seem that protests have spontaneously erupted against the target of their conspiracy. Although this scenario is fictional, it mimics real-life situations such as the Saint Denis example described above. *Go Viral!* thus helps inoculate people by exposing how videos and images can be taken out of context and used to provoke emotional responses and manipulate people. In a large-scale study (Basol, Roozenbeek, et al., 2021), we found that game players became significantly better at spotting COVID-19 misinformation, became more confident in their ability to do so, and were less willing to share such misinformation with people in their network. These findings were replicated in three separate languages (English, French, and German) and the first two effects remained significant for at least 1 week after playing.

Although previous research on prebunking has shown promising results and interventions have been developed that inoculate individuals against visual disinformation,

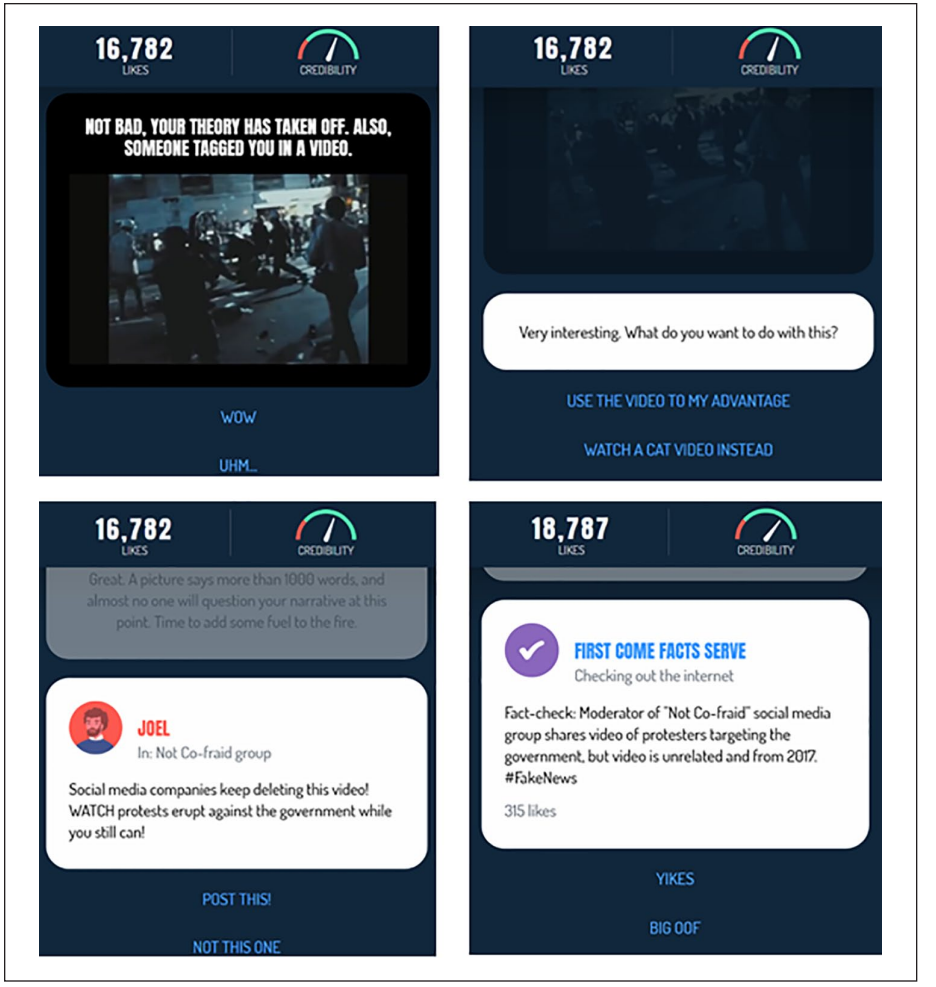

Figure 2. Screenshots from the *Go Viral!* game depicting a scenario with visual disinformation.
Note. "Joel" is the player's character. The blue text at the bottom of the screen are response options that game players can choose between. The top of the screen shows the player's likes and credibility.

we encourage further study on how to measure and combat misleading visual online content, including so-called "deepfakes" (Vaccari & Chadwick, 2020).

Jon Roozenbeek
Postdoctoral Fellow
Department of Psychology, University of Cambridge

Sander van der Linden
Professor of Social Psychology in Society
Department of Psychology, University of Cambridge

Multimodal Misinformation: Next Steps

The contributions of this Invited Forum demonstrate the relevance of multimodal misinformation in modern media environments. In addition, it is clear that research on visual mis- and disinformation is still in its infancy. Nevertheless, it is important to ask what next steps citizens and scholars should take at the present time, considering the effects of misinformation on society.

A good starting point for researchers studying text-based forms of misinformation may be to expand their studies to the domain of *visual* misinformation and disinformation. That is, more research is needed that applies different methodological approaches (e.g., content analytical approaches, laboratory experiments, multiwave panel studies) to study the prevalence and effects of misinformation using both short-term and longitudinal designs.

Given the interdisciplinary nature of this research field (communications, psychology, political science, computer science, etc.), scholars from a diverse set of disciplines should collaborate with one another. Such research collaborations can utilize open science guidelines, and researchers should be encouraged to share both data and their stimulus materials with one another.

Furthermore, it remains unclear whether knowledge about verbal forms of misinformation can generally be directly applied to multimodal formats of misinformation. For instance, do warnings on social media and best practice strategies to debunk text-based misinformation also efficiently work in the context of multimodal information? Is it generally more efficient to use multimodal debunking strategies to correct textual misinformation? (e.g., Hameleers et al., 2020; Young et al., 2018)? And how can continued influence effects of misinformation be best prevented? Future research should systematically test this.

Online platforms like Facebook, Twitter, Instagram, and TikTok must increase efforts to systematically and transparently fight all forms of (visual) misinformation ensuring more democratic and less misleading public discourse. This proposal is increasingly essential as a growing number of citizens around the globe rely on social media for their news. With that said, online platforms need to share more big data sets with researchers (see Hegelich, 2020), for instance, to better understand who spreads (multimodal) misinformation, Also, transparent collaborations between researchers and platforms should be reinforced to fight misinformation.

In addition, research is needed regarding how journalists should respond to the increasing amount of (visual) misinformation. Should journalists working for mainstream media outlets regularly report on and/or correct mis- and disinformation (campaigns; see Tsfati et al., 2020)? If yes, what types of multimodal misinformation should primarily be reported on (with increasing numbers of disinformation campaigns it becomes necessary to select specific falsehoods)? Can reports about this misinformation further divide people regarding their beliefs about what is true and what is false? Could these reports further spread this false content on- and offline? Are journalists able to correctly identify visual misinformation? Future research should explore such questions. In addition, the field of journalism should develop (international) standards and (collaborative) strategies, not only regarding what types

of mis- and disinformation should be covered and corrected, but also how exactly to go about doing this (corrections should provide an alternative explanation, etc.; Lewandowsky et al., 2020).

Furthermore, at a time when our information environments are increasingly polluted with falsehoods and alleged conspiracies, we have to be well prepared to differentiate between correct and trustworthy information and information riddled with falsehoods. Not only should media literacy (i.e., how to spot legitimate and false news; pausing to think before sharing news online; see Fazio, 2020) be taught to children in school, but we must familiarize older populations (who grew up without internet, and rather credible news sources) with how to navigate these new media environments. That is, recent research suggests that news consumers frequently overestimate (i.e., three in four Americans) their ability to correctly distinguish between legitimate information and misinformation (i.e., legitimate and false news headlines; Lyons et al., 2021).

Finally, new forms of inoculating people against visual misinformation should be developed and applied. This means we must further research how individuals can be effectively inoculated against visual misinformation. Then, researchers and journalists can work together to implement inoculation strategies in media coverage—reducing the spread of falsehoods that have the potential to harm democratic societies worldwide.

Christian von Sikorski
Assistant Professor of Political Psychology
Department of Psychology, University of Koblenz-Landau

References

Abel, R. (2004). *Encyclopedia of early cinema*. Taylor & Francis.

Adjer, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). *The state of deepfakes: Landscape, threats, and impact*. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Ahmed, S. (2021). Who inadvertently shares deepfakes? Analyzing the role of political interest, cognitive ability, and social network size. *Telematics and Informatics*, *57*, 101508. https://doi.org/10.1016/j.tele.2020.101508

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trend in the diffusion of misinformation on social media. *Research & Politics*, *6*, 1–8. https://doi.org/10.1177/2053168019848554

Amazeen, M. A., Thorson, E., Muddiman, L., & Graves, L. (2018). Correcting political and consumer misperceptions: The effectiveness and effects of rating scale versus contextual correction formats. *Journalism & Mass Communication Quarterly*, *95*(1), 28–48. https://doi.org/10.1177/10776990166781

Attwood, F. (2007). No money shot? Commerce, pornography and new sex taste cultures. *Sexualities*, *10*, 441–56. https://doi.org/10.1177/1363460707080982

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs*, *77*, 281–311. https://doi.org/10.1080/03637751003758193

Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W., & van der Linden, S. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*. https://doi.org/10.1177/20539517211013868

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about Bad News: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, *3*(1)(2), 1–9. https://doi.org/10.5334/joc.91

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

Berger, C. R., Roloff, M. E., & Roskos-Ewoldsen, D. R. (2010). What is communication science? In C. R. Berger, M. E. Roloff, & D. R. Roskos-Ewoldsen (Eds.), *The handbook of communication science* (pp. 3–20). SAGE.

Bickert, M. (2020, January 7). Enforcing against manipulated media. *About Facebook*. https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/

Breland, A. (2019). The Bizarre and Terrifying Case of the "Deepfake" Video that Helped Bring an African Nation to the Brink. *MotherJones*. https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/

Brennen, J. S., Simon, F. M., & Nielsen, R. K. (2020). Beyond (mis)representation: Visuals in COVID-19 misinformation. *The International Journal of Press/Politics*, *26*, 277–299. https://doi.org/10.1177/1940161220964780

BuzzFeed. (2018). *You won't believe what Obama says in this video!* https://www.youtube.com/watch?v=cQ54GDm1eL0

Cahlan, S. (2020, February 13). How misinformation helped spark an attempted coup in Gabon. *The Washington Post*. https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/

Chesney, R., & Citron, D. K. (2018). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*, 1753–1820. https://doi.org/10.2139/ssrn.3213954

Chesney, R., & Citron, D. K. (2019, January/February). Deepfakes and the new disinformation war. The coming age of post-truth geopolitics. *Foreign Affairs*. https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war

Christopher, N. (2020, February 18). We've just seen the first use of deepfakes in an Indian election campaign. *Vice*. https://www.vice.com/en/article/jgedjb/the-first-use-of-deepfakes-in-indian-election-by-bjp

Citron, D. K. (2016). *Hate crimes in cyberspace* (Reprint ed.). Harvard University Press.

Clarke, Y. D. (2019, June 24). *H.R.3230—116th Congress (2019-2020): Defending each and every person from false appearances by keeping exploitation subject to accountability Act of 2019*. https://www.congress.gov/bill/116th-congress/house-bill/3230

Cole, S. (2017, December 11). AI-assisted fake porn is here and we're all fucked. *Motherboard*. https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn

Compton, J. (2013). Inoculation theory. In J. P. Dillard, & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (2nd ed., pp. 220–236). SAGE. https://doi.org/10.4135/9781452218410

Compton, J., van der Linden, S., Cook, J., & Basol, M. (2021). Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. *Social & Personality Psychology Compass*, *15*, Article e12602. https://doi.org/10.1111/spc3.12602

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowski, S., & Skuce, A. G. (2016). Consensus on consensus: A synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, *11*. https://doi.org/10.1088/1748-9326/11/4/048002

Coopersmith, J. (1998). Pornography, technology and progress. *Icon*, *4*, 94–125.
Cyber Civil Rights Initiative. (2020). *41 states + DC NOW have revenge porn laws*. https://www.cybercivilrights.org/revenge-porn-laws/
Damarla, P. (2018). *Muslim immigrants attack Saint-Denis church in France: Fact check*. www.hoaxorfact.com. http://www.hoaxorfact.com/crime/muslim-immigrants-attack-saint-denis-church-france.html
Dan, V. (2018). A Methodological approach for integrative framing analysis of television news. In P. D'Angelo (Ed.), *Doing news framing analysis II* (pp. 191–220). Routledge.
Daniels, J. (2009). *Cyber racism: White supremacy online and the new attack on civil rights*. Rowman & Littlefield.
Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559. https://doi.org/10.1073/pnas.1517441113
Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics*, *26*(1), 69–91. https://doi.org/10.1177/1940161220944364
Donovan, J. (2021, January). Combating the cacophony of content with librarians. *Global Insights. National Endowment for Democracy*. https://www.ned.org/wp-content/uploads/2021/01/Combating-Cacophony-Content-Librarians-Donovan.pdf
Dreyfuss, E. (2020). Recontextualized media: Biden "voter fraud organization." *Media Manipulation Casebook*. https://mediamanipulation.org/case-studies/recontextualized-media-biden-voter-fraud-organization
Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: A framework and research agenda. *Annals of the International Communication Association*, *43*, 97–116. https://doi.org/10.1080/23808985.2019.1602782
Fazio, L. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Misinformation Review*, *1*(2). https://doi.org/10.37016/mr-2020-009
Franks, M. A. (2017). The desert of the unreal: Inequality in virtual and augmented reality. *University of California, Davis, Law Review*, *51*, 499.
Franks, M. A. (2018). *Beyond "free speech for the White man": Feminism and the first amendment* (SSRN Scholarly Paper ID 3206392). https://papers.ssrn.com/abstract=3206392
Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication*, *37*, 145–156. https://doi.org/10.1080/10584609.2020.1723755
Garcia, M., & Stark, P. (1991). *Eyes on the news*. Poynter Institute for Media Studies.
Garimella, K., & Eckles, D. (2020). Images and misinformation in political groups: Evidence from WhatsApp in India. *Harvard Kennedy School (HKS) Misinformation Review*, *1*. https://doi.org/10.37016/mr-2020-030
Geiger, A. W. (2017, August 30). *Most Americans say libraries can help them find reliable, trustworthy information*. Pew Research Center. https://www.pewresearch.org/fact-tank/2017/08/30/most-americans-especially-millennials-say-libraries-can-help-them-find-reliable-trustworthy-information/
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

Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behavior*, *4*, 472–480. https://doi.org/10.1038/s41562-020-0833-x

Hameleers, M., Powell, T. E., van der Meer, G. L. A., & 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*, 281–301. https://doi.org/10.1080/10584609.2019.1674979

Harding, L. (2012, February). Putin seen behind bars in spoof video. *The Guardian*. https://www.theguardian.com/world/2012/feb/15/putin-behind-bars-spoof-video

Harwell, D. (2019, May). Faked Pelosi videos, slowed to make her appear drunk, spread across social media. *The Washington Post*. https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/

Hegelich, S. (2020). Facebook needs to share more with researchers. *Nature*, *579*. https://doi.org/10.1038/d41586-020-00828-5

Hemsley, J., & Snyder, J. (2018). Dimensions of visual misinformation in the emerging media landscape. In B. G. Southwell, E. A. Thorson, & L. Sheble (Eds.), *Misinformation and mass audiences* (pp. 91–106). University of Texas Press. https://doi.org/10.7560/314555

Iyer, A., Webster, J., Hornsey, M. J., & Vanman, E. J. (2014). Understanding the power of the picture: The effect of image content on emotional and political responses to terrorism. *Journal of Applied Social Psychology*, *44*, 511–521. https://doi.org/10.1111/jasp.2014.44.issue-7

Johnson, N. F., Velásquez, N., Restrepo, N. J., Leahy, R., Gabriel, N., El Oud, S., Zheng, M., Manrique, P., Wuchty, S., & Lupu, Y. (2020). The online competition between pro- and anti-vaccination views. *Nature*, *582*, 230–233. https://doi.org/10.1038/s41586-020-2281-1

Jones, L. K. (2019). *Bringing the receipts: Black feminist theory and intellectual capital in the age of social media*. Available from ProQuest Dissertations and Theses database, 1–182.

Kim, Y., Hsu, J., Neiman, D., Kou, C., Bankston, L., Kim, S. Y., & Raskutt, G. (2018). The stealth media? Groups and targets behind divisive issue campaigns on Facebook. *Political Communication*, *35*(4), 515–541. https://doi.org/10.1080/10584609.2018.1476425

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication*, *50*, 46–70. https://doi.org/10.1111/j.1460-2466.2000.tb02833.x

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Metzger, M. J., . . . , Zittrain, J. L. (2019). The science of fake news. *Science*, *359*, 1094–1096. https://doi.org/10.1126/science.aao2998

Le Pen, M. (2018). *Twitter Post, 19 March 2018*. www.twitter.com. https://twitter.com/mlp_officiel/status/975808872802856960?lang=en

Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., & Zaragoza, M. S. (2020). *The debunking handbook 2020*. https://sks.to/db2020

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*, 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

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

Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 1–38. https://doi.org/10.1080/10463283.2021.1876983

Lyons, B. A., Montgomery, J. M., Guess, A. M., Nyhan, B., & Reifler, J. (2021). Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, Article e2019527118. https://doi.org/10.1073/pnas.2019527118

Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, *27*, 1–16. https://doi.org/https://dx.doi.org/10.1037/xap0000315

Matthes, J., Schmuck, D., & von Sikorski, C. (2021). In the eye of the beholder: A case for the visual hostile media phenomenon. *Communication Research*. https://doi.org/10.1177/00936502211018596

McGlynn, C., Rackley, E., & Houghton, R. (2017). Beyond "revenge porn": The continuum of image-based sexual abuse. *Feminist Legal Studies*, *25*(1), 25–46. https://doi.org/10.1007/s10691-017-9343-2

McGuire, W. J. (1964). Inducing resistance against persuasion: Some Contemporary Approaches. *Advances in Experimental Social Psychology*, *1*, 191–229. https://doi.org/http://dx.doi.org/10.1016/S0065-2601(08)60052-0

McGuire, W. J., & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *Journal of Abnormal and Social Psychology*, *62*(2), 327–337.

Messaris, P. (1997). *Visual persuasion: The role of images in advertising*. SAGE.

Messaris, P., & Abraham, L. (2001). The role of images in framing news stories. In S. D. Reese, O. H. Gandy, & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 215–226). Lawrence Erlbaum.

Nakamura, L. (2002). *Cybertypes: Race, ethnicity, and identity on the Internet* (1st ed.). Routledge.

Nelson, A., & Lewis, J. A. (2019). *Trust your eyes? Deepfakes policy brief*. Center for Strategic and International Studies.

Nisbet, E. C., Mortenson, C., & Li, Q. (2021). The presumed influence of election misinformation on others reduces our own satisfaction with democracy. *Harvard Kennedy School (HKS) Misinformation Review*, *1*(7). https://doi.org/10.37016/mr-2020-59

Noble, S. U. (2013). Google search: Hyper-visibility as a means of rendering black women and girls invisible. *Invisible Culture*, *19*. http://ivc.lib.rochester.edu/google-search-hyper-visibility-as-a-means-of-rendering-black-women-and-girls-invisible/

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism* (1st ed.). New York University Press.

Nyhan, B. (2021). Why the backfire effect does not explain the durability of political misperceptions. *Proceedings of the National Academy of Sciences of the United States of America*, *118*. https://doi.org/10.1073/pnas.1912440117

Nyhan, B., & Reifler, J. (2011). *Opening the political mind? The effects of self-affirmation and graphical information on factual misperceptions*. http://www.dartmouth.edu/~nyhan/opening-political-mind.pdf

Ognyanova, K. (2021). Network approaches to misinformation evaluation and correction. In I. Yanovitsky, & M. Weber (Eds.), *Networks, knowledge brokers, and the public policymaking process* (pp. 1–19). Palgrave Macmillan.

Ognyanova, K., Lazer, D., Robertson, R. E., & Wilson, C. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-024

Paris, B., & Donovan, J. (2020). *Deepfakes and cheapfakes: The manipulation of audio and visual evidence* [Data & Society Report]. https://datasociety.net/library/deepfakes-and-cheap-fakes/

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2015). A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication*, *65*, 997–1017. https://doi.org/10.1111/jcom.12184

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.

Rogers, T., Zeckhauser, R., Gino, F., Norton, M. I., & Schweitzer, M. E. (2017). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of Personality and Social Psychology*, *112*(3), 456–473. https://doi.org/10.1037/pspi0000081

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, *7*(2011199). https://doi.org/10.1098/rsos.201199

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Humanities and Social Sciences Communications*, *5*(65), 1–10. https://doi.org/10.1057/s41599-019-0279-9

Roozenbeek, J., & van der Linden, S. (2020). Breaking Harmony Square: A game that "inoculates" against political misinformation. *The Harvard Kennedy School (HKS) Misinformation Review*, *1*(8). https://doi.org/10.37016/mr-2020-47

Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on "inoculation" theory can reduce susceptibility to misinformation across cultures. *The Harvard Kennedy School (HKS) Misinformation Review*, *1*(2). https://doi.org/10.37016//mr-2020-008

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2018). *FaceForensics: A large-scale video dataset for forgery detection in human faces*. https://arxiv.org/pdf/1803.09179.pdf

Sargent, L. S. (2007). Image effects on selective exposure to computer-mediated news stories. *Computers in Human Behavior*, *23*, 705–726. https://doi.org/10.1016/j.chb.2004.11.005

Sasse, B. (2018, December 21). *Text–S.3805–115th Congress (2017-2018): Malicious Deep Fake Prohibition Act of 2018*. https://www.congress.gov/bill/115th-congress/senate-bill/3805/text

Sawer, P. (2020, December 23). "Deepfake" Queen's Speech: Channel 4 criticised for "disrespectful" Christmas message. *The Telegraph*. https://www.telegraph.co.uk/news/2020/12/23/deepfake-queens-speech-channel-4-criticised-disrespectful-christmas/

Schwartz, O. (2018, November 12). You thought fake news was bad? Deep fakes are where truth goes to die. *The Guardian*. https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth

Snopes. (2018). *Did "Muslim migrants" Attack a catholic church during mass in France?* www.snopes.com. https://www.snopes.com/fact-check/muslim-migrants-attack-catholic-church-mass-france/

Stankiewicz, K. (2019, September 20). "Perfectly real" deepfakes will arrive in 6 months to a year, technology pioneer Hao Li says. *CNBC*. https://www.cnbc.com/2019/09/20/hao-li-perfectly-real-deepfakes-will-arrive-in-6-months-to-a-year.html

Stop Online Violence Against Women. (2018). *Facebook ads that targeted voters centered on black American culture with voter suppression as the end game*. https://stoponlinevaw.com/wp-content/uploads/2018/10/Black-ID-Target-by-Russia-Report-SOVAW.pdf

Stroud, N. J., Thorson, E., & Young, D. G. (2017). Making sense of information and judging its credibility. *Understanding and Addressing the Disinformation Ecosystem*. https://firstdraftnews.org/wp-content/uploads/2018/03/The-Disinformation-Ecosystem-20180207-v4.pdf?x33777

Stubenvoll, M., & Matthes, J. (2021). Why retractions of numerical misinformation fail: The anchoring effect of inaccurate numbers in the news. *Journalism & Mass Communication Quarterly*. Advance online publication. https://doi.org/10.1177/10776990211021800

Sullivan, M. (2019). Libraries and fake news: What's the problem? What's the plan? *Communications in Information Literacy*, *13*(1). https://doi.org/10.15760/comminfolit.2019.13.1.7

Tsfati, Y., Boomgaarden, H. G., Strömbäck, J., Vliegenthart, R., Damstra, A., & Lindgren, E. (2020). Causes and consequences of mainstream media dissemination of fake news: Literature review and synthesis. *Annals of the International Communication Association*, *44*, 157–173. https://doi.org/10.1080/23808985.2020.1759443

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media & Society*, *6*(1), 1–13. https://doi.org/10.1177/2056305120903408

van der Linden, S., & Roozenbeek, J. (2020). Psychological inoculation against fake news. In R. Greifenader, M. Jaffé, E. Newman, & N. Schwarz (Eds.), *The psychology of fake news: Accepting, sharing, and correcting misinformation*. Psychology Press. https://doi.org/10.4324/9780429295379-11

van der Linden, S., Roozenbeek, J., Maertens, R., Basol, M., Kácha, O., Rathje, S., & Steenbuch Traberg, C. (2021). How can psychological science help counter the spread of fake news? *Spanish Journal of Psychology*, *24*, Article e25.

von Sikorski, C. (2021). Visual polarization: Examining the interplay of visual cues and media trust on the evaluation of political candidates. *Journalism*. Advance online publication. https://doi.org/10.1177/1464884920987680

von Sikorski, C., & Knoll, J. (2019). Framing political scandals: Exploring the multimodal effects of isolation cues in scandal news coverage on candidate evaluations and voting intentions. *International Journal of Communication*, *13*, 206–228.

von Sikorski, C., & Ludwig, M. (2018). The effects of visual isolation on the perception of scandalized politicians. *Communications: European Journal of Communication Research*, *43*, 235–257. https://doi.org/10.1515/commun-2017-0054

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

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

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

Wardle, C. (2017, February). Fake news. It's complicated. *First Draft*. https://firstdraftnews.org/articles/fake-news-complicated/

Young, Y. 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*, 49–75. https://doi.org/10.1177/1077699017710453

Zhang, L., & Newman, E. J. (2020). Truthiness: How non-probative photos shape belief. In M. Greifenader, M. Jaffé, E. J. Newman, & N. Schwartz (Eds.), *The psychology of fake news: Accepting, sharing, and correcting misinformation* (pp. 90–114). Psychology Press. https://doi.org/10.4324/9780429295379-11

Zillmann, D., Gibson, R., & Sargent, S. L. (1999). Effects of photographs in news-magazine reports on issue perception. *Media Psychology*, *1*(3), 207–228. https://doi.org/10.1207/s1532785xmep0103_2

Zillmann, D., Knobloch, S., & Yu, H. (2001). Effects of photographs on the selective reading of news reports. *Media Psychology*, *3*(4), 301–324. https://doi.org/10.1207/S1532785XMEP0304_01

Author Biographies

Christian von Sikorski (PhD, University of Vienna) is an assistant professor of political psychology at the Department of Psychology, University of Koblenz-Landau at Landau, Germany. He studies political communication. Specifically, he focuses on political scandals, terrorism, political polarization, and mis- and disinformation.

Viorela Dan (PhD, Free University of Berlin) is Akademische Rätin (postdoctoral researcher) at the Department of Media and Communication of the LMU Munich. Her research focuses on fact checking, the effective correction of mis- and disinformation, with a special focus on rectifying misperceptions resulting from exposure to deepfakes. She is the author of Integrative Framing Analysis. Framing Health Through Words and Visuals (Routledge, 2018).

Britt Paris (PhD in Information Studies, University of California, Los Angeles) is an Assistant Professor in the Department of Library and Information Science at Rutgers University. Paris is a critical informatics scholar studying how groups build, use, and understand information systems according to their values, and how these systems influence evidentiary standards and political action. At Rutgers University, Paris co-organizes the School of Communication & Information Power and Inequality Working Group and the university-wide COVID-19 Communication & Misinformation Working Group with the Institute for Quantitative Biology.

Joan Donovan is a leading public scholar and disinformation researcher, specializing in media manipulation, political movements, critical internet studies, and online extremism. She is the Research Director of the Harvard Kennedy School's Shorenstein Center on Media, Politics and Public Policy, a Lecturer at the Harvard Kennedy School, and the Director of the Technology and Social Change project (TaSC). Through TaSC, Dr. Donovan explores how media manipulation is a means to control public conversation, derail democracy, and disrupt society. TaSC conducts research, develops methods, and facilitates workshops for journalists, policy makers, technologists, and civil society organizations on how to detect, document, and debunk media manipulation campaigns.

Michael Hameleers is Assistant Professor in Political Communication at the Amsterdam School of Communication Research (ASCoR), Amsterdam, The Netherlands. His research interests include framing, populism, disinformation, selective exposure and social identity. He published extensively on the impact of populist communication and misinformation.

Jon Roozenbeek is an ESRC Postdoctoral Fellow in the Department of Psychology and the Cambridge Social Decision-Making Lab at the University of Cambridge. His research focuses broadly on online misinformation, inoculation theory, extremism, and vaccine hesitancy.

Sander van der Linden is Professor of Social Psychology in Society and Director of the Cambridge Social Decision-Making Lab in the Department of Psychology at the University of Cambridge. He co-convenes the Cambridge Special Interest Group on Disinformation and Media Literacy.