Visual disinformation in a digital age: A literature synthesis and research agenda

Teresa Weikmann and Sophie Lecheler
University of Vienna, Austria

Abstract
While a fast-growing body of research is concerned with the detrimental consequences of disinformation for democracy, the role of visuals in this context has so far only been discussed superficially. Visuals are expected to amplify the impact of disinformation, but it is rarely specified how, and what exactly distinguishes them from text. This article is one of the first to treat visual disinformation as its own type of falsehood, arguing that it differs from textual disinformation in its production, processing and effects. We suggest that visual disinformation is determined by varying levels of modal richness and manipulative sophistication. Because manipulated visuals are processed differently on a psychological level, they have unique effects on citizens’ behaviours and attitudes.

Keywords
Communication theory, deepfakes, disinformation, fake images, fake videos, visual disinformation, visual misinformation

Introduction
Few topics have occupied scholars in recent years more than the impact of disinformation on politics and media (e.g. Bennett and Livingston, 2018; Chadwick and Stanyer, 2022; Tandoc et al., 2018; van Aelst et al., 2017). Even though the intentional spread of mediated falsehoods has been evident in political discourse for decades, scholars have declared this the ‘age of disinformation’ (Carmi et al., 2020). Digital information environments have facilitated its extensive, unrestrained and sometimes systematic
distribution (Bradshaw et al., 2020), which is often emanating from populist groups (Hameleers and de Vreese, 2021) and, more recently, has become an elementary component of hybrid warfare (Arcos et al., 2022). Simultaneously, concerns about the trustworthiness of media are rising, while political and mediated discourses about disinformation make it seem like trustworthy information is difficult to find (Allcott and Gentzkow, 2017). In a time when nothing seems believable anymore, a hackneyed phrase comes to mind as a resolution: *‘I’ll believe it when I see it’*. Because what tells the truth if not pictures?

However, surprisingly, the literature on mis- and disinformation has neglected visuals – both as a potential tool to rectify false information and as a dangerous form of disinformation in itself. In particular, debates around ‘fake news’ as pseudo-journalistic news articles and narratives spread around the Internet have rendered disinformation a predominantly textual problem (Wardle and Derakhshan, 2017). While some theoretical studies acknowledge that visuals on their own can constitute misleading content, most papers factor them out of their framework, especially when their primary focus is on fake news (e.g. Egelhofer and Lecheler, 2019; Tandoc et al., 2018). What is more, we know only little about the distinct effects of manipulated visuals (Dan et al., 2021) and even less about who produces them (Thomson et al., 2022). Visual disinformation and its theoretical positioning in today’s media landscape remains a scientific blind spot. This lack of theorising suggests that while research is concerned with the verification of visual disinformation, its impact on democratic processes at large is being generalised (Faulkner et al., 2021). For example, some scholars acknowledge visuals as a complex and challenging form of disinformation (Thomson et al., 2022; Wardle and Derakhshan, 2017), but simplify their expected effect in comparison with text: It is merely argued that visuals amplify disinformation’s impact, but not what concrete attitudinal and behavioural changes can be expected.

We suggest that text and visuals differ in *production, processing* and *effects*: Visuals come in a variety of forms and shapes (Kasra et al., 2018; Thomson et al., 2022), which require certain technological skills depending on their level of sophistication (Kietzmann et al., 2020; Paris and Donovan, 2019). These skills may only be employable by certain actors, which has implications for visual disinformation’s origin and prevalence. In addition, their effect on citizens should be explained through differing models of media effects. First, visuals are processed differently than textual information, creating rich sensory responses that affect our memory (Graber, 1990; Nelson et al., 1976). Second, because they are generally perceived as more credible (Sundar, 2008), and due to their emotionality (Hannah, 2021; Powell et al., 2015), they might have stronger and longer-lasting effects on citizen’s behaviour and attitudes. This means that they can lead to durable misperceptions and increase engagement on social media but also contribute to the erosion of trust in media, politicians and visuals in general (Diakopoulos and Johnson, 2021; Murphy and Flynn, 2022; Puustinen and Seppänen, 2013).

Instead of assuming that visuals only amplify known effects of disinformation, this article is one of the first to conceptualise visual disinformation step-by-step for future research projects. We *first* provide an overview of the various forms through which visuals can disinform, therefore contributing to a better understanding of the scope of disinformation in the digital world. *Second*, we map out the variables at play in the study of
visual disinformation by drawing an effects model. Third, we formulate a research agenda on how to study these effects systematically in order to make a meaningful assessment of their supposed harmfulness.

**Defining and categorising visual disinformation**

In this article, we decide to use the term *disinformation* instead of *misinformation* in order to shift the focus towards the intentionality of manipulating visuals: Deception. So far, scholars have predominantly used the term visual *misinformation* (e.g. Klein, 2020; Shen et al., 2021), therefore leaving aside the malicious intent and deliberate harmfulness of deceptive visuals. Others skip the term completely—simply referring to *fake images* (e.g. Kasra et al., 2018; Singh et al., 2021)—which hampers the integration of visual communication literature into mis- and disinformation research. Disinformation is characterised by malicious intent, but misinformation does not necessarily aim to mislead and can be spread inadvertently (Tsfati et al., 2020). Providing empirical evidence of malicious intent and thus clearly identifying disinformation is difficult in a textual context, as falsehoods can be produced inadvertently (Egelhofer and Lecheler, 2019). However, the case appears to be slightly different for visuals: Constructing false or misleading visual content requires some kind of conscious action and in some instances certain skills. This makes the term disinformation more applicable to the study of visuals. Chadwick and Stanyer (2022) suggest that instead of focusing on the *misperception* caused by false or misleading content, *deception* can function as a bridging concept to put a stronger focus on the origins and intentionality of mediated falsehoods.

**Categories of visual disinformation**

Visual disinformation may be classified along two interdependent dimensions, namely, (1) level of richness in audio-visual modality, that is, if still or moving images are used and (2) level of sophistication, that is, if manipulation occurs using low-level or high-level creation techniques. This categorisation is summarised in Table 1.

<table>
<thead>
<tr>
<th>Modal richness</th>
<th>Low sophistication</th>
<th>High sophistication</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Still images</strong></td>
<td>Elimination, cropping</td>
<td>Photoshopping/doctoring</td>
</tr>
<tr>
<td>Moving images</td>
<td>Changing speed, video filters</td>
<td>Misleading data visualisations</td>
</tr>
<tr>
<td>Both</td>
<td>Decontextualisation</td>
<td>Virtual performance</td>
</tr>
<tr>
<td></td>
<td>‘Cheap fakes’/shallow fakes</td>
<td>Deepfakes</td>
</tr>
</tbody>
</table>

Table 1. Categories of visual disinformation.
disinformation. This can be achieved easily and without even altering the image, for instance, by using a real photograph in combination with manipulated text. In 2016, a real picture of Hillary Clinton accidentally stumbling was used as evidence for her deteriorating health (Cao et al., 2020). Multiple cases of decontextualised images also occurred in connection with COVID-19, where pictures are often used to illustrate or provide evidence for certain false information surrounding the virus (Brennen et al., 2021). Brennen et al. (2021) find that in fact mislabelled, not manipulated images constitute the largest part of COVID-19 misinformation that was identified in fact-checking articles. Such simple ‘decontextualization’ (Hameleers et al., 2020) requires very little technological skills and is therefore low in sophistication (see Table 1). The same also occurs for moving images, where often video footage is simply misattributed with a different date or location. For instance, in 2017, a video of a supposedly drunk migrant assaulting nurses in a French hospital went viral. The video later turned out to show scenes of a Russian hospital with the man in the video being a Russian citizen (Teyssou, 2019).

Manipulation starts when certain parts are removed from a photograph or video, which Kasra et al. (2018) call ‘elimination’. An example stems from Donald Trump’s inauguration in 2017. Here, the photographer was instructed to crop empty space in front of the capitol away to make the crowd appear bigger and busier (Thomson et al., 2022). Minimally edited images are summarised under the umbrella term ‘cheap fakes’, referring to their cost-efficiency. Cheap fakes make use of simple and inexpensive editing techniques such as video filters, speeding up and slowing down or even employ lookalikes. A famous example is a decelerated clip of US politician Nancy Pelosi. It was shared on Donald Trump’s Twitter account with the aim to make her appear drunkenly slurring her speech (Paris and Donovan, 2019).

Even though visual disinformation with a low level of sophistication appears to be more prevalent (Brennen et al., 2021) and is easier to create, it has only been subject to few empirical studies. However, it is possible that textual content analyses of fake news articles have in fact dealt with visual disinformation without explicitly addressing it, for example, if a visual was used to substantiate the claim made in the article.

High sophistication. The emergence of deepfakes has recently given communication scholars the occasion to lay a stronger focus on visual aspects of mis- and disinformation research (Dan et al., 2021). Deepfakes operate on a high level of technological sophistication (see Table 1), as they make use of artificial intelligence to fake someone’s entire audio-visual representation. If both video footage and someone’s voice are artificially generated, Paris and Donovan (2019) speak of a virtual performance, which can be classified as the richest form of visual disinformation (Sundar, 2008). However, there are also examples of deepfake photographs generated with face-swap apps (e.g. ‘Reface’). An often-discussed example in a political context is an educational deepfake video of Barack Obama, in which actor Jordan Peele puts the words ‘Donald Trump is a total and complete dipshit’ in his mouth (Vaccari and Chadwick, 2020). Obama’s face was artificially generated, making it look as if he was really insulting Trump in the clip. Interestingly, the audio was a simple voice imitation in this case and not AI-generated. Though they are not the primary focus of this article, deceptive audios are also a noteworthy form of disinformation today. They especially matter in countries with a preference for
messenger apps like Brazil, where voice messages constitute an important form of information dissemination (Kischinhevsky et al., 2020).

While deepfakes have been taken as an opportunity to test the effects of manipulated visuals on citizens (e.g. Dobber et al., 2021; Lee and Shin, 2021; Vaccari and Chadwick, 2020), research on photoshopped images remains limited. Sophisticated manipulations of still images, also called ‘composition’ or ‘doctoring’ (Hameleers et al., 2020; Kasra et al., 2018), have been possible since the emergence of Photoshop in the 1990s. Back then, they have caused similar concerns about their meaning for the credibility of photographs (Giotta, 2020; Puustinen and Seppänen, 2013). Recent examples are images of natural disasters, such as the Australian bushfire crisis in 2019 (Thomson et al., 2022), where image composition was used to make the disaster scenes look even more dramatic by photoshopping a burning fire in the background of a normal-looking landscape. Similar manipulations also take place in the current COVID-19 crisis, often multimodally in combination with text (Brennen et al., 2021). Even though empirical research has given the prevalence of manipulated still images little attention, it can be assumed that they often occur in combination with textual disinformation (Cao et al., 2020). Highly sophisticated forms like deepfakes, which have triggered serious concerns among journalists (Wahl-Jorgensen and Carlson, 2021), are so far rarely identified in a political context (Bradshaw et al., 2020).

Finally, misleading infographics and data visualisations can constitute visual disinformation, for example, if data are distorted or presented in a way that it either hides or overstates certain relevant parts (Cairo, 2015). In the COVID-19 pandemic, the anti-mask movement has employed tweaked data visualisations to strengthen their argument that concerns about the virus are exaggerated (Lee et al., 2021). Similar actions have been taken by the conspiracy group QAnon (Hannah, 2021). We categorise visual disinformation in form of manipulated data visualisations as sophisticated, as it takes a certain degree of data literacy to manipulate data convincingly. The level of modal richness is low in so far as mostly still images or graphs are used to illustrate certain false claims.

In sum, to categorise something as visual disinformation, it does not matter which kind of visual—still or moving—is used, as long as it is used to accentuate a deceptive message. To do so, an image can either be minimally edited and combined with manipulated text or audio, or completely fabricated to be deceptive on its own. In general, the written or spoken text is providing the false claim, while the visual functions as evidence for it. However, an accompanying text or audio is not strictly necessary. Intentionality is naturally positively correlated with sophistication. While it is possible that a photo was cropped or paired with a wrong caption without malicious intent, artificial intelligence is unlikely to be used unknowingly. To conclude, we define visual disinformation as any information that employs a visual—either in its original form or in a manipulated way—in order to intentionally mislead or deceive, therefore creating an image contrary to reality.1

Production of visual disinformation

Distinguishing between low and high levels of sophistication helps clarifying how visual disinformation is produced and where it originates. While simply adding a false text to a
real image seems like a task anyone can perform, using an algorithm and engaging an actor to fake a politician’s speech (i.e. creating a deepfake) is complicated. Unfortunately, very little is known about the supply side of disinformation in general. However, considering the deceptive strategies and motives of potential suppliers can help enhance our understanding of disinformation’s prevalence (Chadwick and Stanyer, 2022). In this context, Lecheler and Egelhofer (2022) put forward three potential actor groups involved: (1) political actors, (2) media actors and (3) private actors.

Especially deepfakes have recently gained attention as a potentially effective tool for political warfare (Paterson and Hanley, 2020). Yet, they are still very difficult to produce, as their creation requires advanced programming skills (Nguyen et al., 2022). While little is known about malicious actors behind deepfake creation, it appears realistic that high-profile political actors would have the financial means to have them produced on their behalf. For example, the first deepfake created out of a political motivation was by the Indian politician Manoj Tiwar, who commissioned a deepfaked video in which he speaks a Hindi dialect (Pérez Dasilva et al., 2021). There is also evidence for other forms of visual disinformation coming from political actors. To this day, extremist groups and authoritarian regimes regularly make use of manipulated pictorial content to generate followers (Baele et al., 2020). Most recently, decontextualised videos about Russia’s war against Ukraine are perpetually shared on TikTok, with Russian state media being suspected as the driving force behind it (Frenkel, 2022; O’Connor, 2022).

Naturally, this calls into question the role of traditional media, which appears to be dependent on country context. While media practitioners working under authoritarian regimes might engage with disseminating visual disinformation intentionally, places with higher press freedom rather suffer from the accidental spread of visual misinformation. Visuals in itself carry news value (Harcup and O’Neill, 2017), making them likely to be picked up by journalists. Because the detection of false images is sometimes more difficult than for text (Cao et al., 2020; Himma-Kadakas and Ojamets, 2022), the media can drift into the role of accidental disseminators of visual misinformation (Tsfati et al., 2020). Simultaneously, journalists have been found to dominate the discourse about political deepfakes in the news (Wahl-Jorgensen and Carlson, 2021). By stirring up the conversation about deepfakes, they also indirectly contribute to their dissemination and awareness of their existence among citizens.

Finally, financial motives can be identified as reasons for creating disinformation. Visuals are more attention-grabbing than text alone (Geise, 2017), which makes them an attractive asset of fake news stories (Allcott and Gentzkow, 2017). Gaining clicks and other kinds of online engagement through visual disinformation therefore seems to be a plausible motivation behind their creation that private actors may act upon (Tandoc et al., 2018). However, it is difficult to predict the extent of sophistication those might engage in. While it appears unlikely that private actors have the resources to create a convincing deepfake, there are numerous examples of conspiracy groups spreading deceptive visuals online, often in form of misleading data visualisations (Lee et al., 2021). At the same time, visual social media platforms are making it easier to edit, decontextualise and share images quickly (Frenkel, 2022; Wardle and Derakhshan, 2017).

We conclude that while the motives behind creating visual disinformation are not strictly different from textual disinformation, there are different factors to consider. First,
sophisticated visual disinformation is harder to create, which mostly excludes private actors as originators. Second, the media are more likely to fall for visual disinformation, as they are struggling to detect it, which can turn them into disseminators of visual misinformation. Third, visual social media apps diversify the possibilities of all actor groups to create visual disinformation, which suggests a high online prevalence. The provenance and supply of visual disinformation requires its own research focus to make more meaningful claims about their threatening potential, and to find counteracting strategies.

Consequences of visual disinformation

Visual communication literature has a long tradition of investigating the distinct effects visuals have in a news media environment (Coleman, 2010). From this line of research, we know that images are processed differently on a psychological level compared to text, which translates into unique attitudinal and behavioural effects (Geise, 2017; Powell et al., 2015). When connecting this knowledge to the disinformation literature, a number of possible effect paths emerge that are inherent to visual disinformation, which we discuss in the following (see also Figure 1).

Processing of visual disinformation

Two aspects inherent to the processing of visuals distinguish them from text: (1) their high degree of realism, making deceptive visuals increasingly prone to be taken for real and (2) the emotional processing of visuals that may translate into attitudinal and behavioural changes directly.

The faulty processing of information through false visuals (mistaking fake for real). The widespread assumption that images do not lie makes visual disinformation unsettling. This claim finds philosophical justification, as pictures create a sense of reality through their
indexicality (Peirce, 1986). According to Peirce (1986), images establish a physical connection between a photographed object and its spectator. Messaris and Abraham (2001) adapt this to the role of images in news stories and describe that news visuals’ indexicality makes them an effective tool for framing ideological messages, as their content is questioned less critically by the audience. This stands in connection to the realism heuristic of digital media. Sundar (2008) suggests that a sense of realism gradually increases with the modality and richness of representation: The richer the medium, the more real it appears. This means that a picture is richer than text, whereas a video is richer than a picture, leading up to an increase in perceived social presence and higher credibility judgements (Sundar, 2008).

Perhaps this is why the general public is relatively bad at detecting fake visuals online, as people fail to question their authenticity (Kasra et al., 2018). Even though image forgery is an ancient phenomenon, recent findings show that visuals still tend to be trusted over text in a news environment (Hameleers et al., 2020). For example, a focus group study by Kasra et al. (2018) found that doctored still images posted on social media platforms are generally evaluated as authentic. Moreover, Lee and Shin (2021) show that deepfakes have a higher source-vividness compared to text-only and text-photo fake news, which in turn makes them appear more credible and more likely to be engaged with. We conclude that visual disinformation leads to faulty processing of information based on the very fact that visuals generally appear trustworthy and true-to-life.

The processing of visual disinformation through emotions. Visuals are processed differently than text not only on a cognitive, but also on an emotional level. While psychology literature is minimally integrated into research on visual disinformation, the literature on visual framing offers more robust findings on these processes. According to Powell et al. (2015), images in isolation can create stronger framing effects than text alone and impact people’s opinions and intended behaviour without accompanying text. Emotions have a mediating role in these effects. Compared to exposure to text, being exposed to photographs of conflict can increase people’s sympathy, leading to stronger willingness for policy support (Geise, 2017; Powell et al., 2015).

The most common emotions measured in visual framing effects are arguably sympathy, fear, and anger (Geise, 2017), which can be expected to also play a role in disinformation research. For example, combining aesthetic elements and authoritative information in manipulated data visualisations has been an effective tool employed by the QAnon network to access emotional responses of fear and anger (Hannah, 2021). In addition, Vaccari and Chadwick (2020) used uncertainty as a mediator in their study about deepfakes and their effect on trust in news on social media. Uncertainty is considered a cognitive state, but it is closely connected with anxiety: Feeling uncertain, confused or frustrated when seeking information can translate into affective symptoms of anxiety (Kuhlthau, 1993). Their findings indicate that being exposed to a deepfake made participants uncertain, which in turn lowered their trust in news on social media.

Important to consider is the extent to which people are deceived. It appears plausible that being uncertain about whether a news item is trustworthy or not can lead to feelings of anxiety. Also, anger may be elicited when being exposed to—and believing—a scandalous news story about a politician. A key component is the superiority of visuals over
They provide dominant sensory codes that are interpreted more potently by the human brain (Nelson et al., 1976). Whether visual disinformation increases certain emotions such as anxiety and is an important question. It would make visual disinformation potentially more dangerous than text, as emotions can have direct effects on attitudes and behaviour.

**Primary and secondary effects of visual disinformation**

To study the effects of visual disinformation, we propose distinguishing between primary effects—those that happen instantly and upon first encounter with a piece of visual disinformation—and secondary effects—which will only be observable over time.

**Misperceptions due to visual disinformation.** Connected to images’ general perceived believability is the problem of the misperceptions they may cause and how they can be corrected. From a psychological perspective, it is difficult to correct any false information, as learning about the correction is part of processing the false one (Tsafati et al., 2020). This could prove to be particularly dangerous in a visual context. First, images can lead to better memory recall and issue understanding compared to text (Graber, 1990). Second, moving images have the capacity to overwrite existing memories (Murphy and Flynn, 2022). When people witness a situation and then later are presented with a video that displays the situation differently, they tend to believe the video. This is due to the source monitoring network, consisting of three mechanisms: familiarity, imagery and credibility. Initially, images appear familiar to mental experiences. The imagery then overlaps with the memories we carry in our minds. Therefore, they appear as credible proof that a specific event indeed took place (Nash et al., 2009).

Murphy and Flynn (2022) tested the potential of deepfakes versus text-based disinformation to distort peoples’ memories of public events. The findings are mixed: Only one of their manipulations showing a somewhat realistic event resulted in memory distortions, while the other deepfakes did not turn out to be more impactful than manipulated still images or text. The authors suggest that deepfakes can distort memories only if what they portray is plausible and could potentially happen in reality. Recent studies have shown that manipulated photographs can be counteracted by implementing some kind of warning message, for instance, by a fact-checker (Hameleers et al., 2020; Shen et al., 2021). If fact-checkers are practically capable of verifying visual disinformation is another question that needs to be clarified (Vizoso et al., 2021). Overall, there are no conclusive results regarding potential rebuttals of misperceptions caused by visual disinformation. It is possible that they are harder to correct (Dan et al., 2021), and therefore not only stronger, but longer-lasting.

**Engagement with visual disinformation on social media.** Another reason why visual disinformation is considered problematic is because of its shareability (Wardle and Derakhshan, 2017). We already know that fake news stories are very popular on social media which functions as their main disseminator. They get a lot of attention especially on Facebook, where they are liked and shared perpetually, sometimes more than real news stories (Silverman, 2016). The same should be true for visual disinformation, if not more so, as
Loss of trust in visuals. Digital manipulations of news photographs published in mainstream newspapers have in the past led to debates about the general trustworthiness of pictures (Giotta, 2020; Puustinen and Seppänen, 2013). This concern seems to have been refuelled with deepfakes entering the picture. Similar to manipulated still images, they are now considered an epistemic threat: deepfakes can not only produce false beliefs, but actually hinder citizens from acquiring knowledge. When the possibility of seeing a ‘false positive’ increases, the epistemic value of videos is undermined (Fallis, 2021).

Manohar (2020) draws a parallel between deep faked and other computer-generated images in regard to findings of the uncanny valley effect. This phenomenon suggests that people react with uneasiness when encountering a humanoid-looking robot that looks almost like a real person. Processing and making sense of faces requires complex neural structures. The increase of AI- and computer-generated faces poses a new challenge to the human brain, which has to make a distinction between real and fake (Manohar, 2020). Of course, the premise of such manipulations is to look exactly like the real person they are trying to imitate. Still, the discrepancy between fake information and real-looking footage of a familiar face could have precarious neurobiological implications. Repeatedly learning that what we have just seen did in fact not happen might impact the confidence in our abilities to discern between real and fake.

Loss of trust in news media, politicians and science. More articulated than a disbelief in visuals in general is scholarly concern about a potential loss of trust in news media and political institutions. A lot of textual disinformation literature points towards this outcome, studies dealing with fake news in particular (Egelhofer and Lecheler, 2019; Lazer et al., 2018). Initial findings indicate the negative impact of fake news on attitudes towards news media (van Duyn and Collier, 2019). Egelhofer and Lecheler specifically describe that fake news can damage trust in journalistic entities via the fake news ‘genre’ and ‘label’: The genre constitutes pseudo-journalistic content pretending to be real and trustworthy, therefore diminishing the believability of news media. The label means that journalistic content is being accused of being fake in order to delegitimise it (Egelhofer and Lecheler, 2019).

When it comes to visuals, the danger seems to lie in the assumed credibility of images, which is undermined through visual disinformation. For example, Vaccari and Chadwick (2020) found that uncertainty about the veracity of visuals can erode trust in news on social media. If the credibility established by the visual
is compromised, it can lead to a loss of trust in the medium that disseminated the deceptive visual (Puustinen and Seppänen, 2013). Visual disinformation in the form of deepfakes is considered a serious threat to a politician’s reputation. By making them say or do something they never did, but still credulous enough that it seems possible, they can be attacked directly by creating a believable scandal around them (Diakopoulos and Johnson, 2021). Moreover, manipulated data visualisations about the COVID-19 pandemic have led to a decline of trust in journalists and also scientists. Here, counter-visualisations have been used to challenge scientific information about the pandemic (Lee et al., 2021).

To conclude, textual disinformation is considered dangerous rather because it mimics traditional news media formats. In contrast, fear of visual disinformation is driven by visuality: The fact that reality is being mimicked is concerning, as people might lose their baseline trust in all visual information presented in the news, which can in turn erode trust in visualisations provided by journalistic, political and scientific actors.

**Mistaking real for fake.** Beyond the effects stated above, a further unexpected consequence could be that the mere knowledge that visual disinformation exists might confuse citizens to the extent that they start thinking everything is fake—even if the visual is authentic. This proposition stands in connection to the fake news label as described by Egelhofer and Lecheler (2019): Labelling a news story as ‘fake’ is frequently practised by populist politicians to discredit traditional news media. A potential effect is that people might overestimate the amount false information in their media intake, a phenomenon summarised under the term ‘perceived consumption’ (Hameleers and de Vreese, 2021).

Similar processes may be anticipated for visuals, though they might lead to different outcomes, such as a loss of trust in all visuals, no matter the source or context. For instance, Barari et al. (2021) find that partisans are more likely to mistakenly detect deepfakes by assuming that real clips of favoured politicians were fake. When challenged to discern real videos from deepfakes, motivated reasoning made participants discredit actual videos of favoured politicians portraying them in a negative light as fake. Ternovski et al. (2022) showed participants’ warnings about the existence of deepfakes. This led to confusion, making people mistake real content for fake content. The authors note that making citizens overly aware of deepfakes might backfire, resulting in distrust of all kinds of political videos. These concerns appear to be justified, as journalistic discourse about them is driven by fearful conjectures, though there is little evidence for disruption through deepfakes (Wahl-Jorgensen and Carlson, 2021). We suggest that because visual disinformation exists, content that is real might be taken for fake content, which can lead to similar primary and secondary outcomes.

**Studying visual disinformation: a research agenda**

In the following, we outline a research agenda to study the complex forms and unique effects of visual disinformation. Specifically, we propose four key steps based on the effects model we put forward (Figure 1).
Understanding of prevalence and origin

This article started with an overview of the different categories of visual disinformation, which has not been presented in this form before. The categorisation is based on individual cases of noteworthy visual manipulations that have been discussed in the literature. What visual disinformation looks like predominantly and in which form it is most common is understudied, but crucial to understand the extent of the problem. We suggest that future research should shift its focus towards non-Western countries, as the prevalence of forms may be context-dependent. For instance, the role of TikTok as a platform for mis- and disinformation about the war in Ukraine constitutes a current case of interest for the study of disinformation in form of visuals.

To get more knowledge about provenance and supply, fact-checkers that are specialised in detecting mis- and disinformation may be consulted. Due to their expertise, they can provide insights into the overall appearance of visual disinformation and give an estimate of how frequent and dominant visuals are in a disinformation context. In addition, more computational approaches could help to uncover visual disinformation’s prevalence. Even though content analyses have looked into the characteristics of fake news articles that may include images (e.g. Allcott and Gentzkow, 2017; Silverman, 2016), visuals have rarely been addressed explicitly. This is necessary to investigate whether visual disinformation is indeed not only stronger than textual disinformation, but also further-reaching. Content analytical research on visuals needs to be expanded, as to date, only a handful of studies have given empirical evidence for the occurrence of manipulated images (e.g. Brennen et al., 2021). A digital methods approach combined with manual coding seems to be the way forward here to enable the handling of big sets of visual data.

A significant gap in disinformation literature remains the question of who supplies it. Detection techniques are not evolving rapidly enough—neither for text nor for visuals. Understanding the supply side better can work as a preventive measure, as it supports the identification of malicious content. One way to achieve this is by analysing different deceptive strategies of various forms of disinformation. Instead of asking to what extent the audience was deceived, research should focus on how.

Inclusion of visual-specific literature

We propose that visual disinformation can lead to faulty information processing. The mechanism behind this process is different than for text, as visuals have a heightened perceived credibility (Sundar, 2008). As such, a general loss of trust in visuals might result from repeatedly mistaking fake images for real. Another important question is whether mistaking a real image for fake has similar outcomes. We suggest testing this in an experiment, borrowing from concepts such as the indexicality of images (Peirce, 1986) and the realism heuristic (Sundar, 2008). These theories are visual-specific and can help explain the deceptiveness, but also the caused confusion by manipulated visuals.

Next, the study of visual disinformation should consider the mediating effect of emotions. Here, research should make stronger connections with findings from visual
Weikmann and Lecheler

framing (Coleman, 2010; Geise, 2017; Powell et al., 2015) and psychology (Nelson et al., 1976) that highlight the idiosyncrasy of visuals. More specifically, we suggest that visual disinformation may elicit stronger emotions of anger, anxiety and sympathy, based on their content and the extent to which citizens are deceived. In addition, contrasting different issue scenarios and varying intentions of manipulated images, such as harming an opponent or gaining in popularity, may result in different emotional states, which in turn can have different attitudinal outcomes.

Comparison between and within forms

One limitation of this article is that it draws many of its implications from research on deepfakes, which have recently received much scholarly attention (e.g. Lee and Shin, 2021; Vaccari and Chadwick, 2021). Understandably, as a technological novelty, they constitute a gripping and fascinating research topic. However, this level of scholarly attention is not consistent with the empirical prevalence of deepfakes. Instead, research shows that manipulated images are (still) the most prominent form (Bradshaw et al., 2020; Brennen et al., 2021). Thus, future research should consider various formats of visual disinformation and also include manipulated infographics or data visualisations, which have received little attention so far. Methodologically, we suggest a comparison between and within forms. First, contrasting text with visuals (between forms) is of utmost importance, as so far, the claim that images are more dangerous has not been completely clarified. While some studies find that visual disinformation is indeed more credible than text (Hameleers et al., 2020), it is also noted that it is not necessarily more persuasive (Barari et al., 2021).

Moreover, we suggest a comparison within forms, that is comparing varying level of richness (still images vs videos) and sophistication (decontextualisation vs deepfakes). Still images should be included, as they appear more prevalent are easier to produce. Yet, their deceptiveness over richer forms (videos) has not been clarified. We also suggest including cheap fakes or shallow fakes, to test whether sophistication does indeed matter for deceptiveness, and ultimately harmfulness.

Considering primary and secondary effects of visual disinformation

Finally, our proposition that visual disinformation not merely leads to stronger but different effects than textual disinformation requires testing. First, regarding the primary effects, studies need to investigate whether visuals result in longer-lasting misperceptions and how these might be counteracted. Because visuals can affect our memory (Nash et al., 2009), the misperceptions they cause might be longer-lasting. We suggest that different approaches to correct misperceptions caused by visual disinformation have to be investigated to identify most efficient countermeasures. Similarly, the secondary effects should be tested. A general disbelief in all visuals and a potential loss of trust in politicians, even though a more long-term outcome, could be tested in experiments with repeated exposure. Longitudinal studies could give even more insights into how a decline of confidence in senses and institutions develops over time (e.g. Zimmermann and Kohring, 2020).
Conclusion

In a media landscape where visuals are distributed at an outstanding scale, the way they contribute to misperceptions and a decline of trust in institutions remains one of the most critical and disregarded aspects of mis- and disinformation research (Brennen et al., 2021; Dan et al., 2021). Importantly, deceptive visuals likely have distinct effects on citizens, which is why a clearer differentiation between visual and text is crucial. Accordingly, this study extrapolates the intrinsic features of deceptive visuals, whose effects may not merely be stronger than text, but may also be further-reaching and longer-lasting. Furthermore, we provide a more nuanced understanding of visual disinformation by categorising it alongside the two dimensions sophistication and richness. This enables empirical as well as theoretical comparisons between a broader range of forms. That way, future research may achieve greater balance and comprehensive views on the impact of visual disinformation in the coming years, as the field has lately put a disproportional focus on the effects of deepfakes. Moreover, the testable theoretical model may guide studies in identifying pathways that lead to the effects of exposure to visual disinformation. In particular, we suggest paying close attention to the ways in which visuals are processed as opposed to text, as well as the possibility that increased worries about manipulated visuals may lead to mistaking real images for fake.

The provided avenues for future research as well as our own conceptualisation are, however, limited by a number of gaps in the available theoretical literature: First, much of the cited literature is from a Western context, thereby neglecting what role visual disinformation conceptually and empirically may play in other contexts. On the one hand, this affects the understanding of supply and prevalence brought forward in this article, which absolutely depends on the political system it originates from. While we allude to this in our article, we are not able to give an exhaustive overview of the different forms of visual disinformation in a global context. Moreover, this article cannot go beyond offering propositions about the constitution and effects of different forms of visual disinformation. While we do know some things about the role of deepfakes, other forms such as decontextualised images are understudied. Thus, this article raises multiple questions and offers suggestions, but cannot give definite answers. Nevertheless, we believe that it generates a fuller picture of visual disinformation’s scope and uncovers multiple big research gaps, thus serving as a starting point for a wide range of methodological approaches to this highly relevant, yet underexplored topic.

Authors’ note

All authors have agreed to the submission, and the article is not currently being considered for publication by any other print or electronic journal.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Teresa Weikmann https://orcid.org/0000-0001-7508-9757
Note
1. In this article, we are not discussing misleading visual material in form of memes, which can be effective in reinforcing misperceptions. However, memes carry a different cultural meaning than other visuals and are recognised as ‘units of imitation’, thus are inherently not meant to show things as the really are (Shifman, 2013). Similarly, satire or parody is not included, as they differ from disinformation in intentionality: neither are meant to convince citizens of something that is untrue (Egelhofer and Lecheler, 2019).

References


Hannah MN (2021) A conspiracy of data: Qanon, social media, and information visualization. *Social Media + Society* 7(3).


Wardle C and Derakhshan H (2017) Information disorder: toward an interdisciplinary framework for research and policymaking Information Disorder Toward an interdisciplinary framework for research and policymaking. Available at: www.coe.int


Author biographies

Teresa Weikmann is a PhD Student at the University of Vienna and part of the Political Communication Research Group. In her research, Teresa focuses on visuals in the news and how they affect political attitudes. Her PhD project is centered around visual disinformation and deepfakes in particular.

Sophie Lecheler is Professor of Political Communication at the Department of Communication at the University of Vienna. Her research focuses on digital journalism, emotion, and the impact of technology on politics. She has published in a wide range of peer-reviewed journals, such as Political Communication, Communication Research, Journal of Communication, and Digital Journalism.