Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting

Hameleers, M.

DOI
10.1093/ijpor/edae004

Publication date
2024

Document Version
Final published version

Published in
International Journal of Public Opinion Research

International Journal of Public Opinion Research, **36**, 1–9
https://doi.org/10.1093/ijpor/edae004
Advance access publication 1 March 2024
Article

Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting

Michael Hameleers

Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, The Netherlands

All correspondence concerning this article should be addressed to Michael Hameleers, PhD, University of Amsterdam, Amsterdam School of Communication Research, Nieuwe Achtergracht 166, Amsterdam, The Netherlands. E-mail: m.hameleers@uva.nl

Abstract

Visual disinformation has been regarded as convincing because it strongly resembles reality. Yet, we lack a clear understanding of the effects of different forms of audiovisual disinformation—cheapfakes versus deepfakes. To advance the disinformation literature, this paper reports on the findings of two experiments in which participants were exposed to political cheapfakes and deepfakes, respectively. Our main findings indicate that audiovisual disinformation is not perceived as more credible or believable than the same disinformation in textual format. Importantly, deepfakes are perceived as less credible than cheapfakes with a similar de-legitimizing anti-immigration narrative. Although more research is needed, our findings suggest that less sophisticated modes of deception can be at least as credible as more sophisticated forms of artificial intelligence-driven audiovisual fabrication.

Disinformation—the goal-directed dissemination of deceptive information—has been regarded as a key threat to democracy. Among other things, disinformation can increase polarization, enhance cynicism, or cultivate distrust and uncertainty among citizens (e.g., Bennett & Livingston, 2018; Freelon & Wells, 2020). These concerns are voiced even louder for visual disinformation, and deepfakes in particular (e.g., Dan et al., 2021; Westerlund, 2019). In this paper, we define visual disinformation as intentionally deceptive information for which visual information is decontextualized, manipulated, or fabricated and used as "proof" for false or misleading claims (Weikmann & Lecheler, 2023). In this paper, we focus on visual disinformation in the form of audiovisual content (a manipulated or fabricated video), which can be contrasted to visual disinformation based on still images (i.e., memes, manipulated photos).

Even though deceptive information is often visual, we currently lack systematic research on the effects of visual disinformation (Peng, Lu, & Shen, 2023; Weikmann & Lecheler, 2023). In this paper, we focus on two forms of visual disinformation that are high in modal richness: cheapfakes and deepfakes (Weikmann & Lecheler, 2023). Deepfakes can be understood as video-based disinformation created with artificial intelligence (AI) (Westerlund, 2019). Cheapfakes, in contrast, are forms of visual disinformation for which authentic images or videos are re-contextualized to deliberately alter their meaning (Paris & Donovan, 2020). Cheapfakes require relatively little effort, editing skills, and resources. A cheapfake, for example, can use an existing video of a missile strike but crop the video to remove the original context and re-contextualize the video by claiming that it shows a Ukrainian attack on civilian targets in Russia. Applying these affordances, DeSantis used an AI-generated audio fragment mimicking Donald Trump's voice to delegitimize and attack the former president in a television ad (Isenstadt, 2023), which reveals the actual application of AI-powered disinformation in political campaigns.

Arguably, as video-based disinformation signals a stronger connection to reality than texts, and as recipients may be more likely to accept its honesty due to the realism heuristic offered, visual modes of deception may be difficult to detect whilst being highly convincing (Sundar, Molina, & Cho, 2021). Because visuals are often decontextualized and used as proof for deceptive information (Weikmann & Lecheler, 2023), it is crucial to understand the credibility of such modes of deception. Moreover, as visual disinformation may be much more prominent than textual disinformation online (Peng, Lu, & Shen, 2023), we need to comprehend its effects on public perceptions of credibility and authenticity.

To date, empirical evidence on the severity of audiovisual disinformation's impact on democracy is lacking (e.g., Lee & Shin, 2021; Vaccari & Chadwick, 2020). Overall, deepfakes do not result in misperceptions (Dobber, Metoui, Trilling, Helberger, & de Vreese, 2020; Vaccari & Chadwick, 2020) and are only slightly more credible than textual manipulations (Lee & Shin, 2021). In this setting, we can ask whether the resources and expertise required to make deepfakes is worth the trouble, if no clear persuasive advantage may be achieved by creating synthetic videos from scratch. In light of this, cheapfakes may be a viable alternative for malicious actors that aim to signal authenticity (e.g., Chesney & Citron, 2019). Such forms of decontextualization and editing are currently used more frequently than deepfakes (Brennen, Simon, & Nielsen, 2021). However, considering that affordances in

AI are rapidly improving, deepfakes may become more realistic and dominant in the future, which makes it relevant to understand how they are perceived at this stage.

To better understand the differential effects of deepfakes versus cheapfakes, this paper compares the effects of both modes of audiovisual disinformation. Based on two experiments, we specifically compare the effects of deepfakes and cheapfakes on the perceived credibility of (dis)information. By using the same message, targeted political actor, and context of disinformation, this paper allows for a direct comparison of the impact of deepfakes versus cheapfakes as visual disinformation. In addition, by contrasting both modes of audiovisual disinformation to textual disinformation, we can arrive at an assessment of the relative contribution of visuals to the effects of disinformation.

Deepfakes and Cheapfakes as Visual Disinformation

Disinformation can be understood as the intentional creation or dissemination of deceptive information (Chadwick & Stanyer, 2022; Freelon & Wells, 2020). Disinformation may be motivated by different goals, such as sowing discord in foreign democracies (e.g., Lukito, 2020), increasing cynicism, augmenting societal cleavages in domestic political settings (e.g., Marwick & Lewis, 2017), or making financial profits (e.g., Wardle & Derakhshan, 2017). Disinformation may be both cause and consequence of a fragmented and post-factual information order in which facts have become more relative, partisan, and subject to ideological debates (e.g., Bennett & Livingston, 2018; Van Aelst et al., 2017).

Developments in digital technologies may not only play a role in the dissemination of disinformation but also afford the use of techniques (i.e., AI) that make deceptive content seem authentic and similar to the cue-rich nature of information consumed online. Here, we specifically focus on visual disinformation—referred to as an umbrella term for different types of disinformation that rely on different combinations of textual, visual, or audiovisual modalities to alter, decontextualize or fabricate information with the aim to deceive (see, e.g., Dan et al., 2021; Weikmann & Lecheler, 2023). Although visual disinformation may consist of many different practices, most discussions and concerns revolve around the potentially disruptive impact of deepfakes—synthetic videos created with AI to make real persons say fake things (e.g., Westerlund, 2019).

Deepfakes are created with the use of generative adversarial networks (GANs) (e.g., Yang, Ding, Chen, & Li, 2021). GANs operate based on an AI-collaboration of artificial neural networks. More concretely, trained on real footage of the persons depicted in a deepfake (i.e., celebrities or politicians), AI is used to train a model that can realistically generate synthetic footage of the depicted person (Westerlund, 2019). Although the software to create deepfakes is rapidly improving and freely available, it still requires high-skilled labor and/or financial resources and time to create realistic deepfakes (Dobber, Metoui, Trilling, Helberger, & de Vreese, 2020). In addition, extant research has revealed that the effects of deepfakes are not as strong as sometimes assumed. Dobber, Metoui, Trilling, Helberger, & de Vreese, (2020) failed to find direct effects of deepfakes on the evaluation of the depicted political candidate, whereas Vaccari & Chadwick's (2020) experimental findings indicate that deepfakes likely result in confusion rather than deception. Based on a more direct comparison between deepfakes and textual disinformation, experimental studies by Lee and Shin (2021) and Hwang, Ryu, and Jeong, (2021) indicate that deepfakes are only slightly more credible than textual disinformation. Yet, techniques in deep learning and AI are improving rapidly and may become less resource intense and more realistic over time. Compared with earlier studies (e.g., Dobber, Metoui, Trilling, Helberger, & de Vreese, 2020), our experiments rely on much more advanced techniques that are capable of creating hyper-realistic videos. Thus, it is important to update empirical research on the effects of deepfakes as the technologies evolve. The stimuli created for this experiment are more elaborate narratives than used in previous studies (e.g., Dobber, Metoui, Trilling, Helberger, & de Vreese, 2020), and contain a more delegitimizing and extremist political message than other studies (e.g., Lee & Shin, 2021). Therefore, the limited effects of previous studies may also be driven by the message chosen for the manipulation.

Considering that it still requires considerable effort and resources to create realistic deepfakes from scratch, and that the effects of deepfakes are (still) not as disruptive as sometimes assumed, it can be argued that audiovisual manipulations that are less labor and cost intensive may also be effective. In line with this, we particularly contrast deepfakes to cheapfakes (also see, e.g., Chesney & Citron, 2019; Paris & Donovan, 2020). Different from deepfakes, cheapfakes are audiovisual modes of disinformation that require only little technological sophistication, time, and skills to create. It is an umbrella term for visual disinformation that uses basic software to manipulate or decontextualize visual information and strategically alter its meaning. It can, for example, consist of cropping visuals and videos, the manipulation of the audio component of a video, or slowing down/speeding up videos (Paris & Donovan, 2019). Cheapfakes are used for the decontextualization of existing visual materials and are thus lower in sophistication and deviation from reality than deepfakes (Weikmann & Lecheler, 2023).

Importantly, anyone with a computer and basic editing skills can create cheapfakes quickly—which makes them potentially more dangerous than sophisticated deepfakes. This can be exemplified by the findings of Brennen, Simon, and Nielsen (2021) on visual disinformation during the coronavirus disease (COVID-19) pandemic: a majority of disinformation used visual cues, but these were all created with simple tools: there was no real sign of deepfakes in their sample. To date, at least, it seems that deepfakes are mostly created for entertainment purposes, blackmailing, and satire (e.g., Westerlund, 2019), whereas there is little trace of deepfakes in the political realm.

Against this backdrop, it is relevant to compare the effects of deepfakes and cheapfakes applied to political issues. In this paper, we more specifically look at a fabricated message on immigration, in which a conservative political actor is made to express a more extreme anti-immigration and nativist message than this politician has ever voiced in real life. Simply put, to mimic the delegitimizing anti-establishment narrative central to many political disinformation campaigns disseminated by the alt-right (e.g., Marwick & Lewis, 2017), we measure people's responses to visual disinformation in which a mainstream political actor is made to look like a right-wing populist.

The Effects of Deepfakes and Cheapfakes

The power of visual disinformation is based on the ability of audiovisual information to offer a direct index of reality (Powell et al., 2018) or a realism heuristic (Sundar, Molina, & Cho, 2021). This means that audio-visual cues in disinformation should offer an index of reality that overrides the systematic processing of arguments (Sundar, Molina, & Cho, 2021). We can explain this considering dual-process models of information processing (e.g., Chaiken & Trope, 1999; Petty & Cacioppo, 1986) and the heuristic-systematic model specifically (Chaiken, 1989). These models postulate that information processing can follow two different routes that differ in the amount of conscious effort devoted to information processing. For the systematic processing route, more effortful modes of thinking are used, for example, as people pay attention to the argumentation of a message and consider all informational input to arrive at judgements (Chaiken, 1989). Heuristic processing, in contrast, relates to more effortless processing that requires less cognitive effort and capacity, for example, as people pay attention to the source cues, modality, presentation, or style of a message.

Visual compared with textual disinformation may stimulate more automatic or heuristic processing as the realism index or heuristic offers an easy-to-use shortcut to assess the veracity and trustworthiness of information (Sundar, 2008): When something is directly shown instead of described to people, it should be likely to be rated as true. Video-based disinformation in particular may overburden people's processing capacity due to the rich mode of presentation (Sundar, Molina, & Cho, Sundar, Molina, & Cho, 2021). Compared to textual modalities, then, cheap- or deepfakes that rely on a cue-rich format should stimulate more heuristic processing, and the realism heuristic offered by the videos should override the motivation to systematically scrutinize the validity of the presented arguments in the message. Thus, as the richer and more direct representation of reality depicted in audiovisual disinformation compared to text signals authenticity and realism, audiovisual disinformation may circumvent the detection of deception among recipients. As people generally show a bias toward the truth in information processing unless suspicion is actively triggered (Levine, 2014), audiovisual disinformation may—irrespective of its form—exploit the truth bias and create an illusion of truthfulness.

Hence, as people are more likely to find information honest and truthful than to doubt its honesty, and as visual disinformation may strategically exploit this bias by avoiding the activation of suspicion, visual disinformation may have a credibility or believability advantage over text. Thus, by offering seemingly realistic visual proof for deceptive statements, visually framed disinformation may offer a more convincing and direct index of reality than textual representations (e.g., Brennen, Simon, & Nielsen, 2021; Weikmann & Lecheler, 2023). The cue-rich nature of visual disinformation may motivate the heuristic processing of stimuli, whereas texts may motivate people to assess the validity of the arguments put forward (Sundar, Molina, & Cho, 2021). This way, suspicion is more likely to be triggered for textual disinformation.

Cheapfakes and deepfakes represent different forms of visual disinformation that vary on the level of sophistication (e.g., Weikmann & Lecheler, 2023). Although both forms of visual disinformation are rich in modality, the outcome of a deepfake is a synthetic and fabricated video, whereas a cheapfake is based on the decontextualization of existing visual materials (e.g., Paris & Donovan, 2019). This may have different ramifications for their relative credibility compared with textual disinformation. Considering that cheapfakes are based on authentic visual materials, deception may be difficult to (automatically) detect. Hence, cheapfakes expose people to decontextualized, but authentic, fragments of visual information (e.g., Brennen, Simon, & Nielsen, 2021). Deepfakes, in contrast, may be more easily detected as the voice and mouth movements do not match closely, or because the generated voice of the targeted actor does not closely match the original (Hameleers, van der Meer, & Dobber, 2022). However, the added advantage of deepfakes compared with cheapfakes is that such synthetic media can be used to directly depict the targeted actor speaking, which may offer a more direct index of reality compared to the more "indirect" depiction of reality afforded by cheapfakes.

The difference between the modes of creation of deepfakes versus cheapfakes may correspond to different processing routes. Specifically, deepfakes are generated with AI and herewith present a synthetic representation of reality. Based on the current state-of-the-art, people who critically process the message could detect the discrepancy between previous speeches of the depicted political actor and the deepfake, for example, as the generated speech may not match the original voice or because the mouth movements do not closely match the speech (e.g., Dobber, Metoui, Trilling, Helberger, & de Vreese, 2020; Hameleers, van der Meer, & Dobber, 2022). Cheapfakes, however, rely on authentic footage taken out of context. The systematic processing of the message may therefore still circumvent the detection of deception. Based on this reasoning, the detection of deception, and therefore the cue to systematically process the message, may be more likely for deepfakes than cheapfakes.

Yet, an alternative processing route can also be proposed. Compared with deepfakes, cheapfakes offer a less direct index of reality as they only offer secondary "proof" of a political speech that did not happen in real life. The seemingly more direct depiction of the political speech of deepfakes can sidestep the critical processing of message arguments more than cheapfakes, given that the "seeing is believing" heuristic is invoked more for deepfakes than cheapfakes (Sundar, Molina, & Cho, 2021).

Based on these alternative explanations, and given that extant research has not compared the differential effects of deepfakes and cheapfakes, we explore the effects of different modes of video-based disinformation without raising a-priori expectations on the relative credibility of both forms of video-based disinformation. In this paper, we do not only compare deepfakes to cheapfakes, but also contrast them to textual disinformation to explore whether there is a persuasive advantage of both forms of audiovisual disinformation. Based on the reviewed literature on the realism heuristic offered by video-based disinformation (e.g., Sundar, Molina, & Cho, 2021), and extant research on the credibility of visual compared to textual disinformation (e.g., Hameleers, Powell, Van Der Meer, & Bos, 2020), we do expect that video-based disinformation is overall perceived as more credible than textual disinformation.

To arrive at a baseline understanding of the extent to which deceptive information is rated as less credible than truthful information, we contrasted all disinformation

narratives to factually accurate information (control conditions). This helps us to contextualize the differences in credibility between modes of disinformation considering people's veracity assessments related to trustworthy and factually accurate information: Do people differentiate between disinformation and accurate information in the first place?

Following previous effect studies on disinformation (e.g., Schaewitz, Kluck, Klösters, & Krämer, 2020), we look at the perceived credibility/believability of the disinformation messages as the central dependent variable. We introduce the following research questions and hypothesis that compare the credibility of cheapfakes to deepfakes contrasted to factually accurate and authentic information.

RQ_1: What is the difference in the perceived credibility and believability between factually accurate information and disinformation across modalities?

RQ_2: What is the difference in the perceived credibility and believability between deepfakes and cheapfakes?

H1: Deepfakes and cheapfakes are perceived as more credible than textual disinformation.

Methods

Two separate experiments with a between-subjects design were conducted: One experiment contained the deepfake manipulation and one experiment contained the cheapfake manipulation. We use different experiments because the different techniques involved in the creation of a deepfake versus a cheapfake resulted in different contexts for the deceptive speech. Hence, although we could directly manipulate the actual speech of the politician for the deepfake (i.e., the politician himself directly uttered the speech in the video), we could only indirectly associate the deceptive narrative with the politician in the cheapfake condition. More specifically, although the AI affordances used for the deepfake manipulation allowed us to make it seem as if the politician held a made-up political speech, the lower resources used for the cheapfake only made it possible to associate the speech with the political actor (i.e., the speech was repeated in the video by the interviewer, and the politician himself only confirmed that these were his words).

For the cheapfake experiment, we therefore created a different cover story in which we associated the manipulated narrative with an interview allegedly held with the depicted politician. This also meant that the textual and audiovisual control conditions were adjusted to be congruent with the context of an interview setting. This design choice reveals an important difference between the creation of deepfakes versus cheapfakes: Although cheapfakes require less resources to manipulate, the choice of the input material requires more effort and consideration as it must be suited for the strategic de-contextualization of information. We thus rely on two experiments as the deepfake manipulation made the political actor express the anti-immigration message used in this study, whereas the cheapfake manipulation presented this speech as the answer from the politician in an interview. Although the speeches were identical across the studies, different studies were conducted as there were different contexts for the presentation of the disinformation (a political speech versus an interview).

Here, we do need to stress that the data collections were completed in two different periods. Fieldwork for the cheapfake study lasted from November 30 to December 10, 2020, and fieldwork for the deepfake study covered the period from April 1 to April 12, 2021. Yet, as we rely on a low-salient political speech without references to temporal events, as well as a politician who was no longer formally part of the national government for both studies, the different periods are not expected to affect the outcomes. In addition, the anti-immigration sentiments presented in the speech were equally salient and temporal in Dutch society at the time of both data collections. Hence, agreement with the statement that "immigrants are responsible for most of our country's problems" was similar in the cheapfake ($M = 3.81$, $SD = 1.98$) and the deepfake study ($M = 3.75$, $SD = 1.78$). As additional check for comparability, we can see that political preferences were very similar across the two data collections. Most importantly, participants' rating of the political actor depicted in the message was very similar in the deepfake ($M = 36.28$, $SD = 23.87$) and the cheapfake study ($M = 35.42$, $SD = 24.07$).

For both experiments, we used a between-subjects factorial design in which we randomly exposed participants to textual disinformation versus audiovisual disinformation (a cheapfake and a deepfake, respectively). The topic of the disinformation messages was identical: Participants were exposed to a political speech of a conservative Dutch politician (former party leader of the Christian-Democrats). This fragment was about 45 seconds long. In all disinformation conditions, the deceptive narrative was kept constant: The real politician was made to express a fake radical right-wing narrative. Similar to actual practices of manipulation and decontextualization central to disinformation, we altered the meaning and interpretation of an existing speech to make it reflect a delegitimizing and extremist message. The deceptive narrative used in both experiments contained the same claims (see Supplementary Appendix A for the deceptive narrative connecting immigrants to violent crimes). There were no pre-treatment attention checks, and the stimuli were presented after a pre-treatment survey block in which political orientation (a left-right self-placement scale ranging from 1 (left) to 10 (right)) and demographics were measured. After that, participants were randomly allocated to the different treatment or control conditions (equal groups sizes). Finally, they answered the questions for the dependent variables and the manipulation check before they were debriefed.

We consider the stimuli as representing disinformation for different reasons. First, the deceptive and factually inaccurate statement that people from "backwards and retarded societies commit violent crimes" is presented as a "fact." However, empirical research contradicts this claim as crime rates are lower (also proportionally) among illegal immigrants compared to native citizens (see e.g., Light, He, & Robey, 2020). Thus, the statement is factually inaccurate as it is not in line with relevant expert knowledge or empirical facts on crime rates. Second, the statements were never actually voiced by this politician and we deliberately made the politician express more radical viewpoints than can be expected based on his affiliation. This essentially makes the content disinformation as it reveals the intentions of the communicator to deceive recipients (e.g., Chadwick & Stanyer, 2022; Wardle, 2017): The false information on crime rates was deliberately fabricated in visual information to mislead people about crime rates, or to attack and harm an opposed political actor. Third,

we used either de-contextualized visuals (cheapfake) or AI-generated synthetic visuals (deepfake) as "proof" that the deceptive narrative was expressed by the politician—which is a common practice of visual disinformation that can intentionally be used to increase the legitimacy of deceptive claims (see Weikmann & Lecheler, 2023).

Design Study 1: Cheapfake Versus Textual Disinformation

The design of Study 1 concerns a 2 (disinformation: present versus absent) × 2 (modality: textual vs. cheapfake) between-subjects factorial design. The absence of disinformation refers to authentic information that is not manipulated or altered and can therefore be considered as a control group (presented in both a textual and video format). The textual disinformation message presented the deceptive information as a quote resulting from an interview that was recently held with the political actor in question. To make the cheapfake, a real interview with the depicted political actor was manipulated by making it seem as if the interviewer (who was not visible in the video) voiced the deceptive script. By simply recording the audio fragment and presenting it as a summary of what the political actor just said in the interview "So, to summarize the viewpoints you just expressed …" and by then using a strong confirmation sentence of the interviewee in the actual interview "Yes, this is correct." – low effort modes of manipulation were used to put the deceptive narrative in the mouth of the politician.

The audiovisual control condition showed the authentic speech of the political actor, presented as an interview held with the politician. We more specifically used a political speech on immigration and conservative values that was voiced by the depicted political actor. The textual control condition presented the authentic speech of the politician as a quote coming from the same interview. Both the treatment and control thus refer to an interview held with the politician, and indirectly presented the manipulated narrative as if it came from the politician.

Design Study 2: Deepfake Versus Textual Disinformation

The design of Study 2 concerns a 2 (Disinformation: Present versus absent) × 2 (Modality: Textual versus deepfake) between-subjects factorial design. The absence of disinformation refers to authentic information that is not manipulated or altered and can therefore be considered as a control group (presented in both a textual and video format). The deepfake used more sophisticated methods to make it seem as if the political actor expressed the deceptive speech himself. Here, we relied on an extensive collaboration with computer vision and AI-experts to create the deepfake from scratch. A professional voice actor was hired to enact the deceptive speech. Based on GANs and weeks of training, a computer model was trained to create synthetic media of the depicted political actor. As a final result, the mouth and face movements and gestures of the real political actor were matched with the deceptive speech—so that a realistic video of the targeted political actor that closely matched actual footage of this politician was generated.

We contrasted the disinformation messages to an authentic speech of the depicted actor, presented as both a textual and audiovisual message. The same narrative as the control condition in Study 1 was used, but now presented as coming directly from the depicted politician. Hence, because the deepfake approach allowed us to directly alter the political speech of the depicted politician, we did not have to de-contextualize materials in an alleged interview setting.

Sample Study 1

For both experiments, the collection of data was outsourced to the international research company Kantar. This company uses voluntary-opt in panels that are recruited via different methods (e-mail, advertising, telephone). Members of the panel are invited for specific research projects via the company's online invitation system, which also ensures that participants cannot participate in more than one study at the same period. Participants' behavior is closely monitored (i.e., speeding, straight lining) to ensure the quality of responses.

We conducted a-priori power analyses, which indicated that about 60 participants per cell would yield sufficient power (alpha = .05) to detect significant main effects given the small effect sizes previously found in studies on disinformation. Based on light quota (non-interlocked) a diverse sample of participants approaching the Dutch population in terms of age, gender, education, and ideological preferences was collected. For the cheapfake sample, 287 valid completes (out of 350 invites) were used in the analyses. The mean age was 52.72 (SD = 16.61). Of these completes, 30.0% was lower educated, and 27.2% was higher educated (42.9% moderate). Lower education consisted of no formal education/primary school/lower levels of secondary education. Higher education consisted of higher professional education (HBO) and University education (WO) or a PhD/equivalent. Regarding gender, 50.9% identified as female. The sample was also balanced in terms of political orientation: 49.1% identified as (somewhat) left-wing, whereas 50.9% identified as (somewhat) right-wing. Political orientation was normally distributed, with equal proportions in the extreme categories.

Sample Study 2

The distributions of the deepfake experiment's sample are very comparable. In this study, 366 valid completes were retained (out of 450 invites). The mean age of participants was 49.45 (SD = 14.33). Of all complete responses, 25.9% was lower educated, and 32.2% was higher educated (41.9% moderate). For this sample, 54.6% identified as female. Regarding political orientation, 55.5% identified as left-wing, and 44.5% as right-wing. Although the samples across the two studies are quite similar, it could be argued that level of education, ideology, and gender can impact the credibility and believability of disinformation. We therefore also ran the analyses with weights assigned to the responses to make the influence of these factors comparable across samples—these robustness checks yielded the same findings as reported below.

Measures Study 1 and Study 2

For both studies, we measures the believability/perceived credibility of the message that participants were exposed to by using a scale of three separate items that were measured on a 1 (completely disagree) to 7 (completely agree) scale. Only the two extreme scores were labeled (there was no neutral option).

The items were formulated as follows: "The message is truthful," "The message is realistic" and "The message is believable" (Cheapfake sample: M = 3.80, SD = 1.57, Cronbach's alpha = .858; Deepfake sample: M = 3.90, SD =.91, Cronbach's alpha = .805). The averaged scale consisting of the three items was termed perceived credibility. We use a reflective measurement of the underlying construct credibility as we believe that the construct of credibility is the cause of the employed measurement (Edwards, 2011), meaning that the measure of credibility based on the average score of the items reflect credibility or the perceived truth value of disinformation. Similar to other disinformation studies measuring credibility, we also believe that credibility is a one-dimensional construct. The items were developed for this study but inspired by extant experimental research mapping the credibility of disinformation (e.g., Schaewitz, Kluck, Klösters, & Krämer, 2020). We regard this as a measure of the perceived "truth value" of disinformation (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). We averaged the three items as they all mapped participants' evaluation of the extent to which the message was trustworthy and reflective of reality. Additional analyses with the different items inserted as separate dependent variables also yielded the same results. Exploratory factor analyses also confirmed that the items loaded on one factor. Considering that people arguably show a tendency to accept the trustworthiness of incoming information when evaluating it (Levine, 2014), the scores around the midpoint of the scale can be explained in light of the truth bias of information processing.

The credibility measures used in this paper make it difficult to disentangle evaluations of the veracity/credibility of the political statements voiced versus the overall believability of the political message and its authenticity. Although the main aim of this paper was to assess people's overall acceptance of the believability of a political message and differences in these evaluations across forms of disinformation, it is worthwhile to explore the locus of believability. Although our measures do not allow us to clearly differentiate between medium and message, we additionally used a thought listing exercise in both studies. A thematic analysis reveals that believability or credibility was mostly related to the congruence between the source and the message: People who rated the deepfake or the cheapfake as low in credibility mainly referred to the low likelihood that the depicted political actor would express extremist statements, and explained that there was no close fit between the actor's political profile and the message expressed.

Results of Study 1

Table 1 summarizes the mean credibility scores across all conditions. First, answering RQ_1, our findings indicate that there are no significant differences in the perceived credibility/believability across any of the experimental conditions ($F(3, 287) = .14$, $p = .934$, partial $\eta^2 = .002$). More specifically, we can observe a non-significant and small difference between the perceived credibility of textual disinformation (M = 3.87, SD = .99) versus factually accurate textual information (M = 3.88, SD = .84). Likewise, the difference in perceived credibility between a real video (M = 3.95, SD = .80) and a cheapfake is small and insignificant (M = 3.94, SD = .99). Finally, and in contrast to the expectation raised under H1, there is no significant difference in perceived credibility between textual disinformation (M = 3.87, SD = .99) and a cheapfake (M = 3.94, SD = .99). The findings may suggest a truth bias in information processing: Regardless of the veracity and modality of the information, participants tended to regard the alleged interview with the politician as credible.

Table 1. Mean Credibility Ratings Across Experimental Conditions

Condition	Cheapfake		Deepfake	
	M	*SD*	*M*	*SD*
Authentic text	3.88_a	.84	3.99_a	1.28
Authentic video	3.95_a	.80	4.36_a	1.14
Textual disinformation	3.87_a	.99	3.50_b	1.61
Audiovisual disinformation	3.94_a	.99	3.40_b	1.89

Note. Means with differing subscripts between rows differ significantly based on post hoc (Bonferroni) corrected pairwise mean score comparisons.

Results of Study 2

Answering RQ_1, we see a significant difference in perceived credibility across the different conditions ($F(3, 362) = 8.13$, $p < .001$, partial $\eta^2 = .063$). Inspecting the pairwise post hoc (Bonferroni corrected) mean score differences, we first of all see a non-significant difference between textual disinformation (M = 3.50, SD = 1.61) and factually accurate/authentic textual information (M = 3.99, SD = .1.28) – although it can be observed that the factually accurate and authentic message was seen as more believable and credible than disinformation using the same modality ($\Delta M = .49$, $SE = .23$, $p = .213$). Although the difference is not significant, it could be argued that the difference of.49 on a 7-point scale is meaningful, as the factually accurate information was rated as more credible than the disinformation message. Arguably, the non-significant difference is due to the relatively small sample size of the experiment.

The difference in perceived credibility between the authentic video and the deepfake is, however, significant: The deepfake (M = 3.40, SD = 1.89) was perceived as significantly less credible than the authentic video (M = 4.36, SD = 1.14; $\Delta M = .97$, $SE = .22$, $p < .001$). Even more so, the credibility rating of the authentic video was higher than the midpoint of the 7-point scale, which was not the case for any of the other conditions. Finally, there was no significant difference between the perceived credibility of a deepfake (M = 3.40, SD = 1.89) and textual disinformation (M = 3.50, SD = 1.61; $\Delta M = .10$, $SE = .22$, $p = 1.00$). For Study 2, the difference in perceived credibility between an authentic video and a deepfake was more substantial than the difference in the credibility ratings of a cheapfake versus an authentic video in Study 1. Although participants were likely to perceive the authentic video as credible, disinformation was more likely to trigger suspicion in Study 2.

Although the findings indicate that deepfakes and cheapfakes are not perceived as significantly and substantially more credible than textual disinformation, and thus reject H1, we only focused on direct effects in this paper. If we take the congruence between the ideological stance of the disinformation (a radical right-wing message) and people's ideological beliefs into account, it can be observed that people scoring the three highest values on the ten-point ideological self-placement scale overall found the messages more credible (M = 4.15, SD = .98) than people scoring lower on this scale (M = 3.84,

SD = .89). However, there are no significant interaction effects between ideological orientation and exposure to cheapfakes or deepfakes, and the small difference in credibility for left- and right-wing participants are similar for exposure to real information and disinformation.

Conclusion: Are Deepfakes more Credible than Cheapfakes?

To answer RQ_2, we compare the mean scores of the perceived credibility/believability of deepfakes versus cheapfakes. Meeting the assumption that the data is normally distributed, we use an independent samples *t*-test. Based on this test, we find that the cheapfake ($M = 3.94$, $SD = .99$) is perceived as significantly more credible than a deepfake ($M = 3.40$, $SD = .1.89$; $\Delta M = .54$, $SE = .25$, $t(169) = 2.20$, $p = .029$, 95% CI[.06, 1.02]). This finding must be interpreted with caution as the comparison is based on two different data collections. Although the sample, procedure and panel composition are similar, the comparison is less robust as would be the case for a single experimental study.

Discussion

Considering that we know markedly little about the relative believability of deepfakes versus cheapfakes as subtypes of visual disinformation based on video manipulation, this paper compared the credibility and believability of these different types of deceptive information. Although we did not explicitly compare exposure to deepfakes and cheapfakes within a single experimental design, the findings offer first evidence for the responses to different modes of audiovisual disinformation. Contradicting concerns about the disruptive potential of highly realistic deepfakes (Maras & Alexandrou, 2019; Westerlund, 2019), deepfakes were not rated as more believable and credible than the low-cost manipulations of cheapfakes, and substantially less believable than authentic information. Although we have to interpret the findings with care, we even found an indication that—on average—deepfakes are rated as less credible and believable than cheapfakes. Yet, as deepfakes and cheapfakes were both not regarded as highly credible by recipients, the concerns about the unprecedented impact of visual disinformation should not be overstated.

However, it is important to note that textual disinformation and authentic textual information are perceived as equally credible by participants. Considering that the manipulation checks confirm that people perceived the difference in arguments and viewpoints between real information and disinformation, this finding may indicate that people find it difficult to detect deception. This is supportive of extant research on people's (dis)ability to discern true from false information (Luo, Hancock, & Markowitz, 2022) and the idea that the omnipresence of "fake news" accusations have contributed to an information ecology where factual information has become difficult to separate from deception. In an information ecology where different competing truth claims are presented in a cue-rich and seemingly authentic format alongside accusations of disinformation, the distinction between factually accurate information and disinformation may be problematic. As a consequence of uncertainty, people may evaluate information as moderately credible, without clearly discriminating between authentic information and disinformation with a similar ideological slant.

Arguably, in general, cheapfakes could be perceived as relatively credible (at least when compared to authentic information or deepfakes) as there is no fabrication of synthetic media involved in cheapfakes. Cheapfakes are constructed by editing, cropping, and pasting/decontextualizing fragments of a video. Thus, recipients are exposed to the real footage of the depicted actor, whose words are simply placed in a deceptive context by altering the surroundings of a fragment of a real speech. Deepfakes, in contrast, expose recipients to synthetic media created from scratch—which also means that people are listening to a voice actor (or AI-generated speech) and seeing a manipulated face instead of the "real" person that is impersonated. Although deepfakes offer more options for the complete fabrication of deceptive narratives, as they do not rely on the narratives already expressed by the depicted person, they may not be considered as highly realistic.

This paper has a number of implications. First of all, at least when we look at the application of deepfakes to sow discord or amplify cleavages by delegitimizing mainstream political actors using extremist narratives, disinformation may need to use more plausible narratives that more closely reflect familiar perspectives of the depicted politician. Although audiovisual disinformation should correspond to the activation of a realism heuristic (Sundar, 2021), our findings suggest that recipients do not uncritically accept the veracity and honesty of synthetic videos. Thus, even though deepfakes may result in realistic media, news users are relatively critical toward arguments and viewpoints that do not match the profile of political actors. This begs the question of how far disinformation agents can deviate from reality when deceiving recipients, and what the boundary conditions for acceptable manipulations are.

This brings us to limitations and future research suggestions. We compared cheapfakes and deepfakes but only looked at one political actor, one polarizing issue, one degree of facticity, and one country. Future research should explore the differential impact of different types of deceptive content, varying issue positions, degrees of facticity (plausible versus implausible claims) and actors (more or less well-known actors). Future research should also look outside of the political arena, where participants' existing beliefs and knowledge as well as the familiarity of politicians' existing views may dampen the effects of cheap- and deepfakes. Hence, visual disinformation may be more credible for less known and less politicized issue positions. We should also note that we rely on a comparison of deepfakes and cheapfakes across two different studies with relatively small sample sizes. We suggest future research to isolate and control all other factors, rely on a bigger sample, and more directly compare both forms of audiovisual disinformation in one single experimental design. Hence, an explicit limitation is that we did not expose participants to the different forms of audiovisual disinformation within a single experimental design where we could control for unmeasured factors (i.e, the temporal distance, sample composition). Another important limitation is that we look at shorter-term effects on one single dependent variable. The question is whether repeated exposure to deepfakes and cheapfakes can enhance their credibility. Hence, as indicated by Pennycook, Cannon, and Rand's (2018) study, the repetition of disinformation can enhance its believability, which is a consequence of the illusory truth effect (i.e., when the same message is repeated, it is more likely to be regarded as honest as it is similar to messages encountered previously). It is also

important to note that the assessment of credibility did not clearly differentiate between evaluations of the source, content, and presentation of the message (the medium). Insights from a thought-listing task show that evaluations of credibility are likely to be based on the congruence between source and message, and the realism of the presentation. Future research needs to further disentangle the components of credibility evaluations, ideally distinguishing between source, presentation, and content. Finally, based on the alternative processing routes for deepfakes versus cheapfakes proposed in this study, future research may explore systematic versus heuristic processing styles for different modes of audiovisual disinformation.

Despite these limitations, our findings show that different forms of visual disinformation—cheapfakes and deepfakes—are not perceived as more credible than textual disinformation. Although visual disinformation can be considered as a prominent form of deception online (e.g., Weikmann & Lecheler, 2023), we need more evidence on the prevalence of deepfakes and cheapfakes and their indirect effects on media users. At the very least, this study offers tentative evidence that highly sophisticated modes of visual disinformation are not clearly more effective than less sophisticated and more accessible forms of deception.

Supplementary Data

Supplementary data are available at *International Journal of Public Opinion Research* online.

Biographical Notes

Dr. Michael Hameleers (PhD, University of Amsterdam) is an Associate Professor of Political Communication at the Amsterdam School of Communication Research (ASCoR). His research interests include (right-wing) populism, disinformation, and selective exposure.

References

Bennett, L. W., & Livingston, S. (2018). The disinformation order: disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 117–121. doi:10.1177/0267323118760317

Brennen, J. S., Simon, F. M., & Nielsen, R. K. (2021). Beyond (mis) representation: Visuals in COVID-19 misinformation. *The International Journal of Press/Politics*, 26(1), 277–299. doi: 10.1177/1940161220964780

Chadwick, A., & Stanyer, J. (2022). Deception as a bridging concept in the study of disinformation, misinformation, and misperceptions: toward a holistic framework. *Communication Theory*, 32(1), 1–24. doi:10.1093/ct/qtab019

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). The Guilford Press.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. NY, New York: Guilford Press.

Chesney, R., & Citron, D. K. (2019). *Deep fakes: A looming challenge for privacy, democracy, and national security* (California Law Review Research Paper No. 692). Retrieved from https://ssrn.com/abstract=3213954.

Dan, V., Paris, B., Donovan, J., Hameleers, M., Roozenbeek, J., van der Linden, S., & von Sikorski, C. (2021). Visual mis- and disinformation, social media, and democracy. *Journalism & Mass Communication Quarterly*, 98(3), 641–664. doi: 10.1177/10776990211035395

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. H. (2020). Do (microtargeted) deepfakes have real effects on political attitudes? *International Journal of Press/Politics*, 26(1), 69–91. doi: 10.1177/1940161220944364

Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational research methods*, 14(2), 370–388. doi: 10.1177/1094428110378369

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

Hameleers, M., Powell, T. E., Van Der Meer, T. G., & Bos, L. (2020). A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. *Political Communication*, 37(2), 281–301.

Hameleers, M., van der Meer, T. G., & Dobber, T. (2022). You won't believe what they just said! The effects of political deepfakes embedded as vox populi on social media. *Social Media + Society*, 8(3), 20563051221116346. doi: 10.1177/20563051221116346

Hwang, Y., Ryu, J. Y., & Jeong, S. H. (2021). Effects of disinformation using deepfake: the protective effect of media literacy education. *Cyberpsychology, Behavior, and Social Networking*, 24(3), 188–193. doi: 10.1089/cyber.2020.0174

Isenstadt, A. (2023). DeSantis PAC uses AI-generated Trump voice in ad attacking ex-president. *Politico*. https://www.politico.com/news/2023/07/17/desantis-pac-ai-generated-trump-in-ad-00106695

Lee, J., & Shin, S. -Y. (2021). Something that they never said: multimodal disinformation and source vividness in understanding the power of AI-enabled deepfake news. *Media Psychology*, 25, 531–546. doi:10.1080/15213269.2021.2007489

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

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

Light, M. T., He, J., & Robey, J. P. (2020). Comparing crime rates between undocumented immigrants, legal immigrants, and native-born US citizens in Texas. *Proceedings of the National Academy of Sciences of the United States of America*, 117(51), 32340–32347. doi:10.1073/pnas.2014704117

Lukito, J. (2020). Coordinating a multi-platform disinformation campaign: Internet Research Agency activity on three US social media platforms, 2015 to 2017. *Political Communication*, 37(2), 238–255. doi: 10.1080/10584609.2019.1661889

Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility perceptions and detection accuracy of fake news headlines on social media: effects of truth-bias and endorsement cues. *Communication Research*, 49(2), 171–195. doi: 10.1177/0093650220921321

Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos. *The International Journal of Evidence & Proof*, 23(3), 255–262. doi: 10.1177/1365712718807226

Marwick, A. & Lewis, R. (2017). *Media manipulation and disinformation online*. New York, NY: Data & Society Research Institute. https://datasociety.net/output/media-manipulation-and-disinfo-online/

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/

Peng, Y., Lu, Y., & Shen, C. (2023). An agenda for studying credibility perceptions of visual misinformation. *Political Communication*, 40(2), 225–237. doi: 10.1080/10584609.2023.2175398

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental*

Psychology: General, 147(12), 1865–1880. https://psycnet.apa.org/doi/10.1037/xge0000465

Petty, R. E., Cacioppo, J. T., (1986). *The elaboration likelihood model of persuasion* (pp. 1–24). New York: Springer.

Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2018). Video killed the news article? Comparing multimodal framing effects in news videos and articles. *Journal of Broadcasting & Electronic Media*, 62(4), 578–596. doi: 10.1080/08838151.2018.1483935

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

Sundar, S. S. (2008). *The MAIN model: a heuristic approach to understanding technology effects on credibility* (pp. 73–100). Cambridge, MA: MacArthur Foundation Digital Media and Learning Initiative.

Sundar, S. S., Molina, M. D., & Cho, E. (2021). Seeing is believing: is video modality more powerful in spreading fake news via online messaging apps? *Journal of Computer-Mediated Communication*, 26, 301–319. doi:10.1093/jcmc/zmab010

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), 205630512090340–205630512090313. doi: 10.1177/2056305120903408

Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., De Vreese, C., Matthes, J., ..., & Stanyer, J. (2017). Political communication in a high-choice media environment: a challenge for democracy? *Annals of the International Communication Association*, 41(1), 3–27. doi: 10.1080/23808985.2017.1288551

Wardle, C., & Derakhshan, H. (2017). *Information disorder: toward an interdisciplinary framework for research and policymaking* (vol. 27, pp. 1–107). Strasbourg: Council of Europe.

Weikmann, T., & Lecheler, S. (2023). Visual disinformation in a digital age: a literature synthesis and research agenda. *New Media & Society*, 25, 3696–3713. doi:10.1177/14614448221141648

Westerlund, M. (2019). The emergence of deepfake technology: a review. *Technology Innovation Management Review*, 9(11), 39–52. doi: 10.22215/timreview/1282

Yang, C., Ding, L., Chen, Y., & Li, H. (2021). Defending against gan-based deepfake attacks via transformation-aware adversarial faces. In 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China (pp. 1–8). IEEE. doi:10.1109/IJCNN52387.2021.9533868