Deepfakes and the Epistemic Apocalypse

Abstract

It is widely thought that deepfake videos are a significant and unprecedented threat to our epistemic practices. In some writing about deepfakes, manipulated videos appear as the harbingers of an unprecedented *epistemic apocalypse*. In this paper I want to take a critical look at some of the more catastrophic predictions about deepfake videos. I will argue for three claims: 1) that once we recognise the role of social norms in the epistemology of recordings, deepfakes are much less concerning, 2) that the history of photographic manipulation reveals some important precedents for deepfakes which correct claims about the novelty of deepfakes, and 3) that proposed solutions to deepfakes have been overly focused on technological interventions. My overall goal is not so much to argue that deepfakes are not a problem, but to argue that behind concerns around deepfakes lie a more general class of social problems about the organisation of our social practices.

Introduction

> Ours, too, is an age of propaganda. We excel our ancestors only in system and organization: they lied as fluently and as brazenly
>
> C.L.R James, *The Black Jacobins* (1938), p5.

Attention: your beliefs are under threat! Any miscreant with access to a laptop can produce of videos of anyone doing anything they want, using advanced and mysterious forms of deep learning. These *deep fake videos* should spook you out. Here, watch some: Barack Obama saying "Ben Carson is in the sunken place"; Tom Cruise joking about Mikhail Gorbachev and polar bears; Richard Nixon delivering his contingency speech about the destruction of Apollo 11; Elizabeth Windsor delivering an alternative Christmas message.[1] Deepfakes mean that we can no longer trust our eyes; at least when we're looking at videos. Even if deepfakes don't become widespread, their mere possibility is enough to undermine the knowledge we gain from recordings. These videos indicate the start of a new age of epistemic troubles: a *fucked-up dystopia* (Schick 2020), the *infopocalypse* (Ovadya 2018),

[1] For links to these videos see:
https://www.youtube.com/watch?v=cQ54GDm1eL0&t=1s&ab_channel=BuzzFeedVideo
https://www.youtube.com/watch?v=iyiOVUbsPcM&t=24s&ab_channel=Vecanoi
https://www.youtube.com/watch?v=LWLadJFI8Pk&ab_channel=InEventofMoonDisaster
https://www.youtube.com/watch?v=IvY-Abd2FfM&t=109s&ab_channel=Channel4

reality apathy (Warzel 2018), the *collapse of reality itself* (Foer 2018). We're entering *the Epistemic Apocalypse.* This is a technological dystopia, and technology itself is the only solution: we urgently need to invest more in deepfake technology to help us detect fakes.

Although a little hyperbolic, the previous paragraph is a collage of real claims from news coverage and commentary about deepfakes. Let's call the view expressed in this paragraph the *Epistemic Apocalypse* narrative. This narrative began to take shape in 2018 with a series of popular articles: "In the Age of AI is seeing still believing?" (Rothman 2018), "Will Deep-Fake Technology Destroy Democracy?" (Finney Boylan 2018), "the Age of Fake Video Begins" (Foer 2018) and "AI-Assisted Porn is here, and We're All Fucked," (Cole 2017). These were followed by research by legal scholars (Chesney and Citron 2019a), media studies researchers (Paris and Donovan 2019, Vaccari and Chadwick 2020, Shin and Lee 2022), and philosophers (Floridi 2018, Rini 2020, Fallis 2020, Öhman 2020, de Ruiter 2021, Rini and Cohen forthcoming, Matthews forthcoming). At the centre of this narrative are three claims:

1. Deepfakes will have terrible effects on our socio-epistemic practices.
2. Deepfakes are historically unprecedented.
3. The solutions to deepfakes are technological.

The goal of this paper is to take a critical look at these claims. On the first claim, I will argue that the idea that deepfakes mark a significant transformation in the way we gain knowledge from recordings relies on a plausible but incorrect view of the epistemology of recordings. On the second claim, I will show that manipulated recordings have been common throughout history. On the third claim, I will argue that a combination of technochauvinism (Broussard 2018), and the post-truth narrative (Habgood-Coote 2018) has focused attention on technological aspects of the problem posed by deepfakes, to the detriment of the social aspects of this problem.

My goal is not so much to argue that we should be sanguine about the threats of deepfakes, but to point out ways in which the Epistemic Apocalypse narrative has distorted the epistemic problems we actually face. The problem with discourse around deepfakes is that writers are asking the wrong questions, transforming real social problems about enforcing the proper norms of image production and dissemination, and the management of ignorance-producing social practices into hypothetical technological problems about how to detect perfect simulacra. If we focus back on social problems, there remain important reasons to be pessimistic about our epistemic situation, but the primary objects of our concern will be techno-social practices: social practices which are shaped by available technology.

Some points about terminology.

I will use 'interpersonal knowledge' for any kind of knowledge that involves relying on a single person (testimonial knowledge being the central case), and 'personal knowledge' for

any kind of knowledge which involves relying only on oneself (perception, memory, and inference being the central cases). These categories are not exhaustive: a point which will be important below.

I will use 'recording technology' as a general term to refer to any technology that produces vivid reproductions of sensory portions of the world, and 'recordings' to refer to the products of this technology: photographs, videos, sound recordings. I will separate the classification of recordings from their epistemic status. Unlike purists who sharply distinguish between photographs and photomontage (see Lopes 2016) I will use 'photograph' and 'video' to refer to all still and moving images which are the product of photographic technology. Recordings can be accurate or inaccurate, 'straight' or modified.

I will use 'deepfake' to refer to a video which is the product of both a recording technology and a deep learning system. On this usage 'deepfake' includes not only videos which are convincing, non-veridical, and presented with an intention to deceive, but also videos which are unconvincing, veridical, or intended to entertain. For example, a photograph produced by a camera which uses deep learning for its autofocus function will count as a deepfake This is a somewhat unintuitive consequence, but the alternative has similar costs. If we use 'synthetic media' for recordings produced using deep learning, and reserve 'deepfake' for synthetic media that is intended to deceive (Schick 2020: 8-9), none of the examples from the first paragraph will count as deepfakes because they are not intended to deceive. (This consequence goes unnoticed by Schick, who classifies the Obama video as a deepfake (2020: 7)). It bears repeating that deepfakes are only one kind of potentially deceptive recording, and are not necessarily those that we ought to be most concerned about (Paris and Donovan 2019).

1. The Social Epistemology of Recordings

The goal of this section is to argue that the dire predictions about the effects of deepfakes rest on a plausible but mistaken view of the epistemology of recordings. The clearest articulation of the threat of deepfakes is given by Regina Rini in her paper *Deepfakes and the Epistemic Backstop* (Rini 2018). Rini draws on work on the epistemology of photography to argue that at present, videos provide a kind of personal knowledge that doesn't involve interpersonal reliance. She argues that the threat of deepfakes creates a need to rely on the producer of a video to have produced an accurate video, simultaneously transforming videos into an interpersonal form of knowledge and undermining the role of videos as an epistemic backstop for other practices of interpersonal reliance such as testimony. Drawing on work by Dominic Lopes and Sandy Goldberg, I will argue that the literature in the epistemology of photography has employed a false contrast between personal and interpersonal knowledge, and suggest that we might see knowledge from recordings as involving reliance on a *group* of people participating in a norm-governed practice.

Let's start off by fixing in on the epistemic threat element of the epistemic apocalypse narrative. Here's Franklin Foer writing in *the Atlantic*:

> The problem isn't just the proliferation of falsehoods. Fabricated videos will create new and understandable suspicions about everything we watch. [...] In other words, manipulated video will ultimately destroy faith in our strongest remaining tether to the idea of common reality. (Foer 2018)

Like many writers in this field, Foer suggests that deepfake videos will create worries about knowledge from videos, with grave social consequences. This passage raises two questions: how are concerns about deepfakes different from generic sceptical worries? And why is knowledge from videos particularly socially important? If Foer is simply deploying sceptical arguments, standard anti-sceptical responses can be used in response. If he doesn't have a story about why videos are "our strongest remaining tether to the idea of common reality", there isn't a clear reason why we should care more about videos than any other source of knowledge.

Rini argues that videos are a distinctive source of information because they provide a source of perceptual knowledge which doesn't involve interpersonal reliance. On her view, deepfakes threaten a qualitative transformation in the character of knowledge from videos. In a situation in which deepfakes are present, she claims that we will need to rely on the videographer to have made an accurate video, replacing personal with interpersonal knowledge. This qualitative transformation undermines what she takes to be a central social role for videos: providing an epistemic backstop for our testimonial practices. The idea is that videos provide both acute correction by allowing us to check the content of testimony and play a regulative role by creating an incentive to speak truthfully (2018: 2-3). In Rini's view only a personal source of knowledge can provide the backstop for an interpersonal source of knowledge, because a proper backstop needs to have independent epistemic credentials (2018: 10).[2]

The important feature of Rini's view which allows it to vindicate the epistemic apocalypse narrative is the claim that deepfakes lead to a qualitative shift in the epistemic status of videos. As a contrast consider Don Fallis's *The Epistemic Threat of Deepfakes* (Fallis 2018). Fallis argues that the consequence of deepfakes is that videos will shift from an epistemic source with especially high informational value—a view inspired by Meskins and Cohen's view of the epistemology of photography (Cohen and Meskins 2004)—to one with lower informational value. This is a merely quantitative change, meaning that this diagnosis doesn't really support the idea that we are losing touch with common reality.

---

[2] Here Rini's thought seems to be that as a matter of historical contingency videos have come to play a foundational role in regulating our testimonial practices. It is not obvious to me that we haven't rather established a coherentist system of regulation, whereby we calibrate different sources of testimony against one another.. If we can use different testimonial sources to backstop one another holistically, it would not be such a disaster if deepfakes transformed videos into a merely testimonial source of knowledge (see Harris 2021).

The claim that videos provide us with a source of perceptual knowledge is a striking one. To support it, Rini relies on a tradition in philosophical thinking about photographs which starts with the contrast between *photographs* and *handmade pictures* (Sontag 1977: 154, Benjamin 1931/1999: 510, 517-8, 1935/2008, Walton 1984, Cohen and Meskin 2004, Hopkins 2012, Cavedon-Taylor 2013). There are two important differences between photographs and handmade pictures. First, a handmade picture might involve misrepresentations introduced deliberately by the draughtsperson. A photographer can choose which scenes she photographs when but cannot choose to introduce new features into a scene at will (except by changing the scene: think of posed photographs).[3] Secondly, forming beliefs based on the contents of a handmade picture involves relying on the draughtsperson's competence and sincerity, whereas one need not rely on a photographer's competence or sincerity when we form beliefs based on the contents of her photographs. We can gain knowledge about features which are depicted in a photograph which the photographer didn't notice (think of a photobomber in the background of a photograph) (Cavedon-Taylor, forthcoming).[4]

These differences suggest that the distinction between handmade pictures and photographs tracks the distinction between interpersonal and personal knowledge (see Moran 2005). We might think of handmade pictures as something like a visual form of testimony, meaning that when we form a belief based on a picture, our beliefs are mediated by the draughtsperson's beliefs about the scene (Lopes 2016: 20),[5] we extend trust to the draughtsperson's their artistic competence and sincerity and we share the responsibility for the belief formed with the draughtsperson (Cavedon-Taylor 2013). By contrast, the way we gain beliefs from photographs doesn't appear to involve mediation via the photographer's belief, trust in her sincerity and competence, or any kind of shared responsibility. We can simply look at the photograph and form beliefs based on what it shows. Since, the quasi-perceptual phenomenology associated with photographs suggests that they do not provide us with inferential knowledge (Walton 1984, Cavedon-Taylor 2013),[6] we might think that photography must be a *basic* source of personal knowledge (Walton 1984: 263-5).[7]

---

[3] Although the discretion concerning what is depicted cannot introduce misrepresentations, it can generate misleading inferences. A photograph of a group that leaves one person out of the edge of the frame could easily support the inference that they were not present in the scene. See footnote 7.

[4] We do rely on the competence and sincerity of some videographers: the sincerity and competence of a documentary film-maker may well be at issue when we assess whether to trust the narrative of a film (see, Fraser 2021).

[5] We might caveat this claim to deal with creationist teacher-style cases for drawings (Lackey 2008).

[6] Besides perception and inference, we might think of photographs as involving a kind of technologically extended memorial knowledge.

[7] The view that photography doesn't involve mediation by any belief needs to be restricted in scope to beliefs about the features which are depicted in a photograph. We form beliefs based on inferences from the features that we can see in photographs: seeing a picture of a happy couple in wedding attire, we infer that they had a happy wedding day (see Cavedon-

There are several possible non-interpersonal accounts of the epistemology of photography: perhaps photographic knowledge is perceptual (Cavedon-Taylor 2013), involves relying on a mechanical process (Walton 1984), or involves relying on nature itself (Talbot 1844). Rini extends Cavedon-Taylor's perceptual account to the case of videos, arguing that the phenomenology of videos suggests that—at least pre-fakery—videos provide us with perceptual knowledge (2018: 10). Once we become aware of the possibility of deepfakes, when we form beliefs based on videos, we must either extend our trust to the videographer, making videographic knowledge akin to knowledge from testimony, or rely on background beliefs about the likelihood of faking, making it into a kind of inferential knowledge.[8] Either way, videographic knowledge loses its distinctive character as non-interpersonal knowledge which is suitable to play a role in the epistemic backstop.[9]

The contrast between handmade pictures and photographs warms us up to the division of sources of information into personal and interpersonal sources. However, this contrast is a false one: we often gain knowledge by relying on not on individuals, but on *groups* of people. In *Four Arts of Photography* Dominic Lopes puts dependence on photographic practice at the centre of his account of the epistemology of photography. Commenting on the possibility of manipulating digital photographs, he says:

> Despite all the fretting about the danger of digital technology to photography's epistemic credentials, there has been no catastrophe. The reason is not that the technology makes manipulation hard, for it does not. Nor is it that film is still used to ensure the honesty of the signal. On the contrary, film is now more likely to be used in aesthetically oriented practices and has largely disappeared from the newsroom and the forensic lab. Rather. The reason why we continue to trust photography is that, in epistemic photographic practices, photo-manipulation is unprofessional, and is punished. (Lopes 2016: 110).

According to Lopes's view, the epistemic value of a photograph is not secured by the photographer alone, but by a practice of taking photographs which is appropriately

---

Taylor 2013: 292-3). Both photographs and videos are also typically accompanied by testimonial information in the form of captions, narration, and supporting text—think of a family member narrating a photo album—and many of the beliefs we glean from recordings rely both on captions and the features depicted in a photograph. Exactly how to sift out the content depicted by a recording from that which is accessible in light of background knowledge and supporting testimony is a difficult issue.

[8] This is not an exclusive or: a reductionist about testimony will think of testimony as derivative of knowledge from inference.

[9] This means that from Rini's point of view, responses which appeal to our ability to rank the reliability of different sources of videos (see Harris 2021) are effectively conceding that deepfakes have transformed the epistemic status of videos.

regulated by a defensible set of social norms of photographic practice.[10]

In a series of papers Sandy Goldberg has developed a general social dependence view of the epistemology of instruments—thermometers, watches, weighing scales—into which we can situate the epistemology of recordings as a special case. Echoing the contrast between handmade pictures and photographs, Goldberg focuses on the distinction between testimonial and instrumental knowledge to investigate the distinctiveness of testimonial belief (Goldberg 2012, 2020). He observes that testimonial beliefs involve reliance on a particular individual, who can be held responsible for normative failures, and who receives partial credit for true beliefs. By contrast, beliefs based on an instrument seem to involve reliance on an object, which cannot be held responsible, and can receive no credit for true beliefs. In the earlier paper Goldberg leans towards a personal picture of the epistemology of instruments (Goldberg 2012). He later refines his position, stressing the fact that instruments are artefacts: objects which have been designed to track features of the world, whose reliability depends on the way they are designed, operated, and maintained (Goldberg 2020). Goldberg argues that while beliefs we gain from instruments do not involve reliance on a particular person, they do manifest a kind of *diffuse* epistemic dependence.[11] In a case of diffuse epistemic reliance, we base our beliefs on the proper operation of a social practice (see Goldberg 2011a). For example, when we employ reasoning of the form 'if that were true, I would have heard about it by now', we do not rely on any particular individual. Rather we rely on a set of information-dissemination practices—newspapers, gossip, public information broadcasting—to provide us with timely information on a question (Goldberg 2011b). Goldberg suggests that when we rely on an instrument, we rely on the social practices relating to the production, operation, and maintenance of that instrument. If these practices are governed by appropriate norms, then beliefs formed based on that instrument will be justified, and if these norms are not met, beliefs based on that instrument (even when true) will be unjustified. Although we cannot blame the instrument, Goldberg claims that we have practice-generated entitlements to hold producers, operators, and maintainers to the norms of the relevant social practices.

If we think of the epistemology of recordings as involving diffuse epistemic reliance on the social practices of operating recording technology, processing recordings, and disseminating them to viewers, the epistemic status of a video isn't dependent on a particular videographer; it depends on whether the relevant social practices are being effectively regulated by appropriate norms. There is a substantial question to be asked about what the appropriate norms for operating recording technology (Morris 2020: 84-9), for

[10] If reliance on a social practice is central to the epistemology of photographs, there need not be a qualitative difference between photographs and hand-drawn pictures. Hand-drawn pictures can be the product of a well-regulated social practices: consider archaeologists' drawings of artefacts, and ornithologists' drawings of bird species (Cohen and Meskin 200, 17, Lopes 2016, 87-91, 108-13).

[11] For Goldberg's views on the relations between epistemic norms and social practices, see (Goldberg 2018)

editing videos (Meek 2019), and for presenting videos to an audience (Atencia-Linares 2012) are, but I take it as given that producing inaccurate deepfakes and disseminating them as real videos is a violation of the norms of producing and disseminating videos. This means that the existence of deceptive deepfake videos will downgrade the epistemic standing of beliefs based on videographic practices. However, the baseline is not the kind of reliability we might associate with perception. There are already several important ways in which the appropriate norms of videography are being broken. Videos are edited in misleading ways (Meek 2019), are presented out of context (Reuters 2022), and are manipulated using a variety of techniques, including simply changing their speed (Reuters 2020). From the perspective of this view of the epistemology of recordings, the existence of deepfakes is a quantitative downgrade in the epistemic status of beliefs which are based on a social practice which already involves some flouting of important norms.

Goldberg doesn't provide us with an account of the epistemic status of beliefs which are based on practices that involve partial compliance with appropriate norms, but I think that the reasonable view would be that we can gain some important justification for beliefs which are based on practices that involve non-catastrophic norm flouting. The supporters of the epistemic apocalypse narrative might combine their view with a diffuse reliance picture of the epistemology of videography, and contend that deepfakes are in fact a catastrophic norm violation, rendering the practice of making and disseminating videos unsuitable for epistemic reliance. Part of the job of the next section is to argue that this view would collapse into a general scepticism about knowledge from recordings, given how widespread deceptive manipulation has been, historically speaking.

2. The History of Manipulated Recordings

An important part of warnings about the imminent Epistemic Apocalypse is the idea that deepfakes are historically unprecedented.[12] This is a particular worry if we think—with Rini—that before the apocalypse, our knowledge from recordings was a kind of personal knowledge, but the supporter of the diffuse reliance view might still think that manipulated views present is a significant downgrade in the epistemic status of videos. Here's Schick expressing the view that deepfakes are unprecedented:

> Until relatively recently, the manipulation of media – photos, video and audio – was the domain of specialists or those with immense resources, like a national government or a Hollywood studio. Technology is making human manipulation of media easier and more accessible to everyone. But now, AI has granted humans a new tool by giving machines the power to generate wholly synthetic (or fake) media. This technology is still nascent, but we are in the early stages of an AI revolution

[12] Some writers do acknowledge the existence of manipulated photographs and videos (see Rini 2020: 11-12, Matthews forthcoming), but these examples are presented as isolated exceptions

which will completely transform representations of reality through media. (Schick 2020 25-6)

In part, this impression is created by the magic of neologisms, which encourages to infer that a new word refers to a wholly new kind of thing. (We might call this error the *neosemantic fallacy.*[13]) This section has three goals: to correct the historical record by showing how common manipulated photographs have been, to consider how manipulated photographs in the news have been handled, focusing on the widespread fakes in US photographic media from 1880 to the 1920s, and to suggest that the harms of failures to abide by proper norms of photographic practice have accrued to racialised minorities.

Most histories of photography focus on technological innovations, take 'straight' photography as the central case, and prioritise photography's epistemic rigour over its aesthetic possibilities. In *Faking it: Manipulated Photography before Photoshop,* Mia Fineman develops a counterhistory which turns the standard narrative on its head. Fineman focuses on the way in which manipulation was used to accommodate technological limitations, discusses several traditions which were organised around systematic manipulation, and details the aesthetic, propagandistic, and playful uses of photographic technology.[14]

Fineman starts at the very beginning of photography in the 1840s (Fineman 2012: C1). Early photography had considerable technical limitations, which photographers accommodated through a combination of skill with the technology and tampering with their negatives and prints. The differences in light levels and the way that negatives reacted with blue light made it difficult to capture the details of a scene without overexposing the sky, so many 19th century outdoor photographers created composite photographs which combined images of the sky and earth. In many cases photographers re-used the same negative of the sky in multiple pictures. Portrait photographers enriched monochrome prints by employing painters to colorise their pictures, blurring the line between photography and handmade pictures. Although we're accustomed to thinking about retouched photos as a new phenomenon, by the 1850s retouched photos were being displayed in exhibitions, sometimes along with the original photos. Many portrait studios employed artists to beautify their clients. These practices were sufficiently prevalent that in the nineteenth century photographic societies in London and Paris tried to prevent retouched photos from being displayed at their exhibitions.

Trick photographs were popular as postcards, featuring decapitated figures, enormous livestock, and fantastical romantic scenes. Some of these tricks were overt: it is unlikely that anyone was convinced by photographs of a hunter shooting elephant-sized rabbits in Iowa

[13] Thanks to Thi Nguyen for this label.

[14] This book is based on a 2012 exhibition at MoMA, which was sponsored by Adobe and the Andrew W. Mellon foundation.

(Fineman 2012: 142). Others hid their manipulation, and presented their composite prints as real photographs. Famously, William Mumler's spirit photographs—produced by multiple exposures—were published as real photographs, and seem to have convinced a significant portion of the public. Following Mumler's trial, the *New York World* published a column with striking similarities to the apocalyptic predictions about deepfakes we saw in the first paragraph:

> Who, henceforth, can trust the accuracy of a photograph? Heretofore, we have been led to believe that nature, the whole of nature, and nothing but nature, could be "took"; but now whither shall we turn when it is possible for Henry Ward Beecher, say, to be presented I the embraces of a festive *Fleurette,* or the ghost of the late lamented to be delineated with a rawhide in the hand hectoring a gang of negroes in a cotton field? What ravage will this possibility make of private reputation, and what confusion entail on the historian of future times. Photographs have been treasured in a belief that, like figures, they could not lie, but here is a revelation that they may be made to lie with a most deceiving exactness. (*New York World* [*sic*], May 4, 1869, 8, quoted in Tucher 2022 94-5)

Similar trick techniques found their way into early moving pictures: in 1900 Georges Méliès released a short film entitled *L'Homme Orchestre,* in which he used multiple exposures to create the illusion that seven versions of him were playing instruments together. Footage of the Spanish-American war in 1898 involved widespread fakery, including staged cavalry charges, and the use of toy boats create footage representing naval engagements (Tucher 2022: 105-17)

The pictorialist movement, which was dominant in early photography, included many photographers who not only engaged in retouching and other 'enobling' processes, defending these processes by appealing to aesthetic values. Many figures in this tradition refused the distinction between the choices made when taking a picture before exposing a negative, and those made in changing a negative after exposure. In a notorious article in *Camera Work* from 1903, Edward Steichen argues that "every photograph is a fake from start to finish, a purely impersonal unmanipulated photograph being practically impossible" (Steichen 1903). This article exemplifies one stage in an ongoing debate about the proper role of manipulation within aesthetically-oriented photography.

Photographic faking often took place in politically charged contexts. In 1871, in the aftermath of the Paris Commune, the portrait photographer Eugène Appert produced a series of images entitled *Les Crimes de la Commun* (see English 1983). These photographs depicted atrocities carried out by the communards, including the imprisonment of women and priests, and several executions by firing squad. The photographs were labelled with dates and a list of the people depicted. Although presented as genuine photographs, these images were sophisticated montages. Working after the fall of the Commune, Appert took background shots of the streets he wanted to represent, separately took posed shots of

actors to represent the protagonists, replacing their heads with those of communards, and stuck the negatives of these protagonists against the scene, finishing off with artful painting and retouching to create the impression of a genuine photograph. This series was widely sold as prints and postcards and were politically inflammatory. It is not clear whether these pictures were taken to be genuine, but it seems likely that a French public uninformed about photographic tricks take would have taken them seriously; indeed, in the 1930s a historian took them to be real photographs of staged events (English 1983 fn7).

Figure 1. Massacre des dominicains d'Arcueil route d'italie no.38 le 26 mai 1871 à 4 heures et demie, photomontage produced by Eugène Appert 1871. (Image is published by the MET, under an open access public domain copyright, see https://www.metmuseum.org/art/collection/search/701997, and is also available through the Wikimedia foundation)

Appert's photomontages of the Commune are not an isolated case of manipulated photographs being presented as documentary evidence. From the 1840s until the 1880s, the majority of photographs printed in illustrated newspapers were reproductions of photographs made by hand, many with embellishments. When half-tone printing allowed photographs to be printed rather than reproduced, printers in the United States and Europe maintained their habit of tweaking their pictures by hand. In 1884, Stephen Horgan claimed in *Photographic News* that:

> Very rarely will a subject be photographed with the composition, arrangement, light and share, of a quality possessing sufficient 'spirit' for publication in *facsimile.* All photographs are altered to a greater or lesser degree before presentation in the newspaper. (Horgan 1884: 427-8, quoted in Fineman 2012, 140)

In 1898, an editor of a photography magazine summed things up more pithily:

> Everybody 'fakes' (Welford 1898: 572, quoted in Tucher 2017: 206).

The production of photographs during this period has been described as *semi*-mechanical due to the extent of human intervention in the production of images (Beegan 2008: 177). Some of these changes were presumably innocuous—colour corrections, and cropping to create effective compositions—but there was also a tradition of printing composite pictures with an extremely shaky connection to reality. *The New York Evening Graphic* (known as the *porno*-graphic) notoriously produced staged photos and 'composographs' to illustrate current events, often printing sexualised pictures of women (Fineman 2012: 144). In 1902, in an address to the Photographers' Association of America, a photographer defended the practice saying: "I believe in faking, I admire legitimate faking, faking that produces the results desired," arguing that faking allowed photographers to overcome the "falseness of ultra-realism" to attain "not literal, but *spiritual* and *eternal* truth (Parkinson 1902, quoted in Tucher 2017: 198).

Figure 2: Composgraph illustrating the scandal of Peaches and 'Daddy' Browning captioned 'Mad as a Scene from the House of Usher', printed in New York Evening Graph

28th January 1927 (available https://commons.wikimedia.org/wiki/File:Peaches_%26_Daddy_Browning_Composograph_1.gif)

How did newspaper photography transition from widespread manipulation, into the golden era of documentary photography in which manipulation was rare and frowned upon? The historian Andie Tucher argues that debates about faking in photography followed the pattern established in earlier disputes over the status of written fakes (Tucher 2017, 2022). Introducing false details to add 'colour' to stories was common in journalistic writing in the 1880s, but by the late 1890s journalists had established a distinctive professional identity which involved a commitment to reproducing facts, and 'fake' had shifted from a neutral term to describe a journalistic technique to a professional term of derogation (Tucher 2017: 197). Tucher shows how photographers working for newspapers deployed a combination of an appeal to the mechanical objectivity of the camera, and their public commitment to truth-seeking to establish trust in their pictures, and similarly deployed 'faker' against photographers who manipulated (Tucher 2017: 208-10). This is not to say that manipulated photos were unknown—the Graphic's *composographs* were printed in the 1920s and 1930s—but the development of the professional identity of the documentary photographer did establish a practice of photography in which photographers were both trusted and trustworthy, within which manipulated photos counted as norm violations.

After the 1930s, photographic manipulation was perhaps less prevalent, although it found uses in art photography, in propaganda, and satire (Fineman 2012: C3, C6). The introduction of digital cameras, then photoshop made it easier to amateurs to manipulate their photographs, although Fineman points out that many of the affordances of this software reproduce much older techniques (2012: C6).[15]

Fineman's history focuses on the manipulation of photographs through editing and creative techniques, but there is also an important story to be told about how the calibration of photographic technology itself has historical systematically misrepresented parts of the social world. Until the 1980s, photographic film and development processes were set up to accurately depict the colour of white skin, and the colour reference charts used by printers featured white women with pale skin (known as 'Shirleys' after the first model) (Roth 2009, 2019). These choices in the design of photographic technology, and the prescribed methods of film processing meant that photographs systematically misrepresented the appearance of people with non-white skin, changing the tone of their skin, often quite literally hiding the faces of people with dark skin in artificial shadows. Although this kind of oppressive

[15] Digital retouching of photographs precipitated a panic around trust in photography, recycling themes which emerged after Mumler's fakes. In 1985 Whole Earth Review published an issue with a manipulated photograph of a UFO on the cover with the subtitle: 'DIGITAL RETOUCHING: the end of photography as evidence of anything', and in 2008 Fred Ritchin published a book entitled *After Photography* (Ritchin 2008).

technology—in which representational choices are baked into the representational process (see Liao and Huebner 2021)—is importantly different to recordings which are manipulated by hand, once we recognise the epistemic significance of social practices around the design and operation of instruments, we can see that both kinds of systematic misrepresentation are failures in the social norms around the design and operation of photography.

Shifting our attention from the technological magic of the camera to the social practices around the production, processing, and dissemination of recordings, we can see that the idea of a golden period in which recordings unproblematically accurately represented the world is a fiction. Of course, we shouldn't replace an over-optimistic history with a generalised pessimism: the point is that the history of recordings is a mixed bag, involving both reliable and unreliable social practices.

There are four points to take away from this section. First, that manipulation and unreliable depiction have been widespread in the history of photography, perhaps prevalent in some periods and genres. This means that any attempt to claim that deepfakes are a catastrophic norm violation will entail a general scepticism about knowledge from recordings. Secondly, rather than being historically unprecedented, there are uncanny precedents for the various uses of deepfakes: sexualised composographs anticipate deepfake pornography, Appert's faked photographs anticipate the use of deepfakes in political propaganda (with faceswaps replacing headswaps), and trick photography anticipates the satirical and playful use of deepfakes. Thirdly, besides intentional manipulation, recordings have systematically misrepresented in virtue of the way that photographic technology was designed and operated. And fourthly, manipulation has been addressed as much by changing social practices around the production and reception of recordings as by changing recording technology.

3. The Politics of The Epistemic Apocalypse

So far, we've seen reason to be sceptical of the idea that deepfake videos would lead to a qualitative change in the epistemic status of videos and established that deepfake videos are not as unprecedented as proponents of the epistemic apocalypse would have us think.
In this section, we turn to the solutions proposed by supporters of this narrative. We will see how the epistemic apocalypse narrative is shaped by the technochauvanist tendency to centre technological solutions (Broussard 2018) with the distorted history and conservative politics of the post-truth narrative (Habgood-Coote 2018). Diagnosing these narratives both undermines some of the key ideas of the epistemic apocalypse narrative and points us towards a view of deepfakes as a social problem.

3.1. Technochauvinism

As I will use the term, *technochauvinism* is an intellectual attitude, involving three

tendencies: to repackage social problems as technological problems (*techno-solutionism*), to believe that technological systems can perform complex tasks (*techno-optimism*), and to ignore or underplay the importance of the designers, operators, and maintainers of technological systems (*techno-fixation*). We can see this narrative at work across the contemporary public sphere, from discourse about self-driving cars (the solution to our traffic woes, that's been just around the corner for ten years or so despite manifest technical issues, but don't ask about the people in the global south working in exploitative conditions to produce data-sets to train image recognition algorithms) to blockchain technology (the solution to post-2008 problems with trust in banking, that will soon be able to carry out immensely complex calculations without colossal carbon emissions, but don't pay attention to the interests of the people who are designing and running exchanges and issuing non-fungible tokens). These three intellectual dispositions are not intrinsically bad; they are bad because they have bad epistemic consequences for peoples' beliefs about political problems, what technological systems can do, and the role of people within technological systems.

Having identified the basic ingredients of technochauvinism, we can unpack how they are at work in discourse around deepfakes.

Start with techno-solutionism. The idea that deepfakes can be ameliorated by the development of more technological tools is a persistent theme throughout the literature on deepfakes (see Chesney and Citron 2019: 1787, Rini 2018: 7, Schick 2020: 195-8). This can seem obvious: once we've packaged up deepfakes as technological problem, it is natural to think that the solution will come in the form of a technological innovation. Techno-solutionism functions as a focal device: by directing our attention toward the production of new forms of technology whilst ignoring the social conditions which enables this technology to have harmful effects, it obscures the ways in which technologies interact with social practices, and makes it appear that our political agency can only be manifested through the development of new technologies. In the case of deepfakes, directing our attention towards uncanny manipulated videos distrust us from the social conditions of widespread (and justified) distrust in media institutions, and political systems. Rather than thinking about media reform or institutional political change, we end up thinking about how best to detect deepfakes. It is possible that effective deepfake detection technology might be developed, but it is not obvious either that this technology will solve the problems raised by deepfakes or that technological detection will work better than skilled human detection.[16]

Techno-optimism manifests in the way we think about what deepfakes are. Professionally produced deepfakes are unrepresentative of the genre, and most are superbly janky. During the Russian invasion of Ukraine in 2022, deepfake videos of Vlodymyr Zelensky and Vladimir Putin surfaced, both apparently part of genuine propaganda efforts (Wakefield 2022). Neither is remotely believable: Zelensky's head sits off-centre from his neck, the video of his

[16] Comparisons of automated and human detection of deepfakes use untrained human subjects, see (Groh et al. 2022).

face is at a different resolution to the rest of the video, and the skin on his face is a different colour to his neck. Putin's face is similarly pixelated, his expressions are unnatural, and his teeth pop in and out of the video. In May of 2022 an apparent deepfake video of Elon Musk surfaced, advertising a cryptocurrency scam (if you will excuse the pleonasm) (Lawrence 2022). This video is extremely weird: Musk and his interlocutor appear to be voiced by a text reader, their words are out of sync with the video, and Musk's eyes and head move mechanically back and forth. Those who are concerned about the Epistemic Apocalypse correctly point out the fact that currently existing videos are unconvincing doesn't entail that undetectable deepfakes are not just around the corner: Rini describes her paper as an exercise in *prophylactic* political epistemology. It is possible that the technology to produce perfect simulacra of public figures might soon become available, but actually existing deepfakes show us how difficult such videos are to produce.

The techno-fixation of the deepfake discourse comes out in myths about how deepfakes are produced. Researchers suggest that making a deepfake is a matter of getting hold of a video of the target and pressing a button.[17] The reality is rather different: deepfakes require large image sets of the target being edited in (which is why movie stars are favoured subject), knowledge about the quirks of the software, awareness of common problems, and considerable computing power.[18] It is notable that the most effective uses of the technology have relied on skilled impersonators who already have more than a passing resemblance to the targets (Makuch 2021), and combine deep learning with traditional visual effects techniques (Mui 2021). Although it's nice to give the people who create and feature in deepfakes due credit, the main reason for highlighting the skill required to produce realistic deepfakes is to inform our sense of possible interventions.

Identifying technochauvanism both allows us both to debunk false claims about technology by showing that they are unsupported by the evidence, and to reframe the problems we face as social and political problems. Deepfakes are overall less realistic than we have been encouraged to think, and substantially rely on human skills, including old-school impersonation. Although there is a possibility that technological solutions might ameliorate the problem of deepfakes, it is not obvious that they are the only or best solution, and we should also consider social solutions.

3.2. The Post-Truth Narrative

The epistemic apocalypse narrative also relies on what we might call the post-truth narrative (Habgood-Coote 2018). The post-truth narrative was developed after 2016 as an

[17] The minimisation of human agency in the description of the production of deepfakes is in interesting tension with the idea that part of the epistemic threat of deepfakes arises from the improper intrusion of human agency into the production of videos.
[18] For a non-technical description of how a beginner would go about creating a deepfake, see (Lee 2019).

attempt to make sense of the political situation post-Brexit and the election of Donald Trump, and is given its clearest expression in a slew of popular books (D'Ancona 2017; Davies 2017, Ball 2017; McIntyre 2018, Schick 2020). This narrative presents several epistemically worrying phenomena—political lies, troll factories, deepfakes—as symptoms of a relatively recent epistemic crisis, whereby the political culture or institutions of Western democracies have failed to live up to enlightenment epistemic norms (Habgood-Coote 2018: 1054-5). This crisis is to be resolved by a reinvigoration of the enlightenment epistemic norms, which principally manifests in a defence of establishment knowledge-generating institutions (major newspapers, broadcasting corporations, fact-checkers), which proponents of this narrative present as bastions of enlightenment values. If anything, the discourse around deepfakes has heightened the tenor of this narrative, transforming a crisis into an apocalypse.

Habgood-Coote (2018) and Finlayson (2019) discuss several reasons to be sceptical of the post-truth narrative: i) the failure of 'post-truth' to have a determinate meaning, ii) the false history propounded by proponents of the post-truth narrative, and iii) the conversative politics of proposed solutions.

First, it is not clear what the phrase 'the post-truth era' picks out, making it difficult to understand exactly what is supposed to be distinctive about this putative period. 'Post-truth' might refer to a period characterised by widespread bullshit, alternative epistemologies, political beliefs which have lost contact with reality, a loss of truth or the belief in truth, or the failure to value truth (Habgood-Coote 2018: 1043, Finlayson 2019). As things stand neither expert use nor general linguistic dispositions appear to fix the meaning of this phrase between these possible meanings. There is a real risk that the phrase 'the post-truth era' is nonsense, and sentences which use it fail to express determinate meaning.[19]

Secondly, this narrative presupposes a false history in which institutions were unproblematically organised around enlightenment epistemic values. It doesn't take a lot of historical inquiry to show that this is false. The history of propaganda, white supremacy, and European Imperial projects provide us with rich examples of institutions organised around the production of ignorance (Mills 2007).

Thirdly, the narrative encourages a kind of political conservatism which favours the defence of establishment institutions over other kinds of interventions (Habgood-Coote 2018: 1054-8, Finlayson 2019). If the problem is that we have moved from the truth era to the post-truth era, our task is to embrace tradition and to defend establishment epistemic institutions. This conservatism manifests in a focus on newspaper funding and establishing

[19] If 'post-truth' is nonsense, how can we talk about the post-truth narrative? Although the claim 'we are in a post-truth era' may be nonsense, we can still mention the phrase, allowing us to say that people assert that we are in a post-truth era are committed to an overlapping set of claims that do not rely on nonsense phrases.

trust in politicians in the post-truth discourse. Although there might be something to be said for defensive measures, exclusively focusing on them distracts us from interventions which aimed at underlying problems with knowledge-generating institutions.

Similar problems of linguistic failure, bad history, and political conservatism show up in the deepfake discourse.

Although it is not particularly common to explicitly use the phrase 'post-truth' in discussions of deepfakes (Chesney and Citron 2019b), we find similarly vague and underspecified terms sprinkled throughout the discourse. See: 'fucked-up dystopia' (Schick 2020), 'the infopocalypse' (Ovadya 2018), and 'reality apathy' (Warzel 2018). These terms encourage us to think in very general terms about our epistemic predicament, avoiding specific or precise evaluation of practices or institutions. We encouraged to think that the question at issue is *are things fundamentally broken, or basically fine?* and to treat all evidence of epistemic dysfunction as supporting a general pessimism about the present. This attitude can easily shade into nihilism: if our epistemic culture is fundamentally broken, we might think that it is simply beyond recovery.

Proponents of the Epistemic Apocalypse narrative propound a false history of recordings, according to which the widespread malicious manipulation of recordings is a relatively recent phenomenon. We have seen that this history is false: the manipulation of recordings was historically widespread, and in some contexts was prevalent. Forgetting the history of photographic manipulation both encourages us to think of deepfakes as a novel problem, and amplifies our perception of the seriousness of the problem.

There is also an important thread of political conservatism in commentary on deepfakes. In the final chapter of *Deep Fakes and the Infocalypse,* Nina Schick (2020) discusses several parts of a response to deepfakes: i) raising awareness about misinformation and deepfakes, ii) supporting credible journalism and fact-checking organisations, iii) developing technical tools for detecting deepfakes, and authenticating reliable information, and iv) developing institutions for counteracting political misinformation. Revealingly, she describes ii) and iii) as *defence* strategies. These interventions are not themselves bad; the problem is that Schick's focus is on shoring up 'establishment' sources of journalism, failing to reckon with the underlying problems with contemporary journalism.

The post-truth narrative supports two important elements of the epistemic apocalypse narrative: it encourages us to think that historically recordings have been an impeccable source of information, and it encourages us to focus on solutions which defend the status quo. We've already seen in section 2 that there is a long history of manipulated recordings, and there is no reason to focus entirely on defence strategies, to the exclusion of interventions which address the underlying causes of epistemic dysfunction.

3.3. Deepfakes are a Social Problem

What happens if after identifying the technochauvinist impulse and the post-truth narrative, we set both to one side? Building on the suggestion from section 1 that videos involve reliance on a social practice, I want to suggest that we see deepfakes as a social problem. Rather than seeing deepfakes as a problem about a uniquely dangerous form of technology indicative of a fall from a prelapsarian state of epistemic grace, we should see the existence of deepfakes as a symptom of long-running problems around the management of the norms of producing and disseminating recordings.

This social problem does have an important technological dimension. The affordances and organisation of a technology affects the social practices which emerge around it, just as the wider social practices in which a technology is embedded effect how it works. The practices which emerge around technological systems are properly speaking techno-social: effected by both the form of a technology, and the wider social and political context in which that technology functions. The affordances of contemporary deepfake technology do allow the possibility of producing manipulated videos, but whether videos are created (and who they depict) is a matter of the social context in which this technology is deployed.

We have a much better sense of how design better social practices than we do how to nullify problematic forms of technology. Drawing on the example of how early documentary photographers handled faking, we might consider whether a similar combination of linguistic advocacy and community norm policing might lead to a social norm against the dissemination of deepfake videos. Rather than activating semantic associations of '*deep*fake' with 'deep learning', we might instead lean on the associations between 'deep*fakes',* and problematic faking in early photojournalism. Perhaps there is some mileage in attempting to turn 'deepfake' into a term of criticism within a practice of community norm policing. There are some interesting precedents for successful online community norm policing of fakes. For example, in 2014 Shafiqah Hudson and I'Nasah Crockett established the community hashtag #YourSlipIsShowing to call out fake Twitter accounts purporting to be run by Black people (Hampton 2019).

Drawing on our discussion of technofixation above, we might also consider how changes in social practices might affect the practical knowledge required to produce realistic deepfake videos. Although it might be a little optimistic to think that we might uninvent deepfakes by destroying the practical knowledge required to produce convincing videos (Donovan and Paris 2019),[20] removing the financial incentives for producing deepfake pornography, banning forums on mainstream sites, and taking down opensource deepfake tools might be cheaper and more effective interventions for reducing the number of realistic deepfakes in circulation than investing in developing costly detection technology.

[20] But see Mackenzie and Spinardi (1995) on how close the United States got to uninviting nuclear weapons.

It is worth stressing that the most serious harms of deepfake videos are likely to be consequences of established ignorance-producing social practices affecting minority and marginalised groups. If we think about pornography as a kind of propaganda which upholds and maintains a misogynistic social order[21], we can see the threat of pornographic deepfakes is not that they persuade anyone that they are a genuine depiction of the target woman—the current level of technology is far from being up to this task, and we can use background knowledge to infer that famous actors have not starred in pornographic films (see Harris 2021: 13386)—but is rather their ability to effectively spread sexist propaganda. Deepfake pornography presupposes that women are f*ungible* subjects, whose faces and bodies can be swapped around like a children's toy (see Nussbaum 1995). We might worry that the political threat of deepfakes is not that people are persuaded that politicians have said that things that they have not—politicians can always present counter-evidence—but that deepfakes will become an effective tool for white supremacist propaganda (Mills 2007). Virtual influencers like @shudu.gram and @koffi.gram and image generation systems have already demonstrated that software for producing realistic images and videos has considerable potential for commodifying and dehumanising racialised minorities (Jackson 2018, Sobande 2020, Heikkilä 2022). It does not take much imagination to see how deepfake technology might be used to spread controlling images of racialised minorities (Hill Collins 2002). The post-truth narrative encourages us to focus on the epistemic problems of mainstream political discourse. By shaking free of it, we can recognise that the most serious harms of epistemic dysfunction often effect minority and marginalised groups.

The common theme running through our discussion of the epistemology, history, and politics of recordings is that producing and disseminating accurate recordings is a social problem. Presenting deepfakes as a technological problem not only leads us into factual errors concerning the way deepfakes are produced, how realistic they are, and their historical novelty; a technosolutionist and conservative conception of the possible responses to deepfakes ignores the centrality of social practices to the production of recordings and obscures the way in which deepfakes function within more general ignorance-producing practices. Rather than thinking about deepfakes as a science-fiction technological problem, we should think about them as part of our thinking about the actual techno-social practices involved in the production of knowledge and ignorance.

Conclusion

In this paper I have pursued three connected lines of criticism against the view that deepfakes are the harbingers of the Epistemic Apocalypse. First, I showed that construing knowledge from recordings as a special case of knowledge from instruments has the consequence that deepfakes simply make explicit our existing reliance on social practices around the design, operation, and maintenance of recording technology. Secondly, I argued that manipulated recordings have been common historically, and presented an historical

[21] (McGlynn 2016). This idea has its source in the work of Catharine MacKinnon (1987) and has been developed by Rae Langton (1993).

episode in which widespread deceptive photographs were handled via changes in social practices. Thirdly, I argued that a combination of technochauvinism and the post-truth narrative has meant that social measures to address deepfakes have been obscured. The thread running through these three lines of criticism is the idea that if we want to understand the problems posed by deepfakes, we need to think about them not as posing a technological problem, but as a social problem about the management of our practices for producing and receiving recordings.

I anticipate that this paper will be interpreted as proposing a negative answer to the question of whether we should be concerned about deepfakes. Although I have argued that many claims made about the bad effects of deepfakes are false, the goal of this paper is to argue that *are deepfakes bad?* is simply the wrong question. We should be thinking about how to design norms for techno-social practices, including—but not limited to—those involved in the production of recordings.[22]

References

Abrams, Lawrence (2022) Elon Musk deep fakes promote new BitVex cryptocurrency scam, *Bleeping Computer,* May 22nd 2022 https://www.bleepingcomputer.com/news/security/elon-musk-deep-fakes-promote-new-bitvex-cryptocurrency-scam/

Atencia-Linares, Paloma (2012). Fiction, Nonfiction, and Deceptive Photographic Representation. *Journal of Aesthetics and Art Criticism* 70 (1):19-30.

Ball, James. (2017). *Post-truth: How bullshit conquered the world*. Biteback Publishing

Beegan, G. (2008). *The mass image: a social history of photomechanical reproduction in Victorian London* Basingstoke: Palgrave Macmillan.

Benjamin, Walter. (1931/1999). Little history of photography. *Selected writings*, *2*(Part 2), 1931-1934.

Benjamin, Walter. (1935/2008). *The work of art in the age of mechanical reproduction*. Penguin UK.

Broussard, Meredith. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press Cam (MA).

Cavedon-Taylor, Dan (2013). Photographically based knowledge. *Episteme* 10 (3):283-297.

Cavedon-Taylor, Dan (forthcoming). Arrangement and Timing: Photography, Causation and Anti-Empiricist Aesthetics. *Ergo: An Open Access Journal of Philosophy*.

Chesney, B., & Citron, D. (2019a). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, *107*, 1753.

Chesney, R., & Citron, D. (2019b). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs.*, *98*, 147.

[22] Thanks to Natalie Ashton, Thomas Brouwer, Dan Cavedon-Taylor, Haixin Dang, Daniel Elstein, Simon Graf, Nadja El Kassar, Andrew Peet, and Robbie Williams. This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 818633)". .

Cohen, Jonathan & Meskin, Aaron (2004). On the epistemic value of photographs. *Journal of Aesthetics and Art Criticism* 62 (2):197–210.
Cole, Samantha (2017) AI-Assisted Porn is here, and We're All Fucked, *Motherboard,* 12 November 2017 https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn
Cole, Samantha (2018) There is no Tech Solution to Deepfakes *Motherboard* 14th August 2018, https://www.vice.com/en/article/594qx5/there-is-no-tech-solution-to-deepfakes
Collins, P. H. (2002). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. routledge.
d'Ancona, Matthew (2017). Post-truth: The new war on truth and how to fight back. Random House.
Davis, Evan (2017). *Post-truth: Why we have reached peak bullshit and what we can do about it*. Little, Brown Book Group.
de Ruiter, Adrienne (2021). The Distinct Wrong of Deepfakes. *Philosophy and Technology* 34 (4):1311-1332.
English, D. E. (1983). Political photography and the Paris Commune of 1871: the photographs of Eugène Appert. *History of Photography*, *7*(1), 31-42.
Fallis, Don (2020). The Epistemic Threat of Deepfakes. *Philosophy and Technology* 34 (4):623-643.
Fineman, Mia (2012). *Faking it: Manipulated photography before Photoshop*. Metropolitan Museum of Art.
Finlayson, Lorna (2019). What to Do with Post-Truth. *Nordic Wittgenstein Review* 8:63-79.
Finney Boylan, Jennifer (2018) Will Deepfake Technology Destroy Democracy, *the New York Times ,* October 17, 2018 https://www.nytimes.com/2018/10/17/opinion/deep-fake-technology-democracy.html
Floridi, Luciano (2018). Artificial intelligence, deepfakes and a future of ectypes. *Philosophy and Technology* 31 (3):317-321.
Foer, Franklin (2018), The Era of Fake Video Begins, *The Atlantic* May 2018, https://www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877/
Fox Talbot, William Henry, (1844) *The Pencil of Nature,* Longman, Brown, Green, and Longmans, London
Fraser, Rachel (2021). Narrative testimony. *Philosophical Studies* 178 (12):4025-4052.
Goldberg, S. (2011b). If that were true I would have heard about it by now. *Social epistemology: Essential readings*, 92-108.
Goldberg, Sandy (2011a). The Division of Epistemic Labor. *Episteme* 8 (1):112-125.
Goldberg, Sanford C. (2012). Epistemic extendedness, testimony, and the epistemology of instrument-based belief. *Philosophical Explorations* 15 (2):181 - 197.
Goldberg, Sanford C. (2020). Epistemically engineered environments. *Synthese* 197 (7):2783-2802.
Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2022). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences*, *119*(1), e2110013119.
Habgood-Coote, Joshua (2019). Stop Talking about Fake News! *Inquiry: An Interdisciplinary Journal of Philosophy* 62 (9-10):1033-1065.
Hampton, Rachelle (2019) The Black Feminists Who Saw the Alt-Right Threat Coming, *Slate* April 23rd 2019 https://slate.com/technology/2019/04/black-feminists-alt-right-twitter-gamergate.html

Harris, Keith Raymond (2021). Video on demand: what deepfakes do and how they harm. *Synthese* 199 (5-6):13373-13391.
Heikkilä, Melissa (2022) The viral AI avatar app Lensa undressed me—without my consent, *MIT Technology Review,* December 12th 2022 https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/
Hopkins, Robert (2012). Factive Pictorial Experience: What's Special about Photographs? *Noûs* 46 (4):709-731.
Horgan, Stephen (1884) "Photography and the Newspapers," *Photographic News*, 38.1348, 428.
Jackson, Lauren Michele (2018) Shudu Gram Is a White Man's Digital Projection of Real-Life Black Womanhood, *The New Yorker*, May 4, 2018, https://www.newyorker.com/culture/culture-desk/shudu-gram-is-a-white-mans-digital-projection-of-real-life-black-womanhood
James, C. L. R. (2001/1938). *The Black Jacobins: Toussaint L'ouverture and the San Domingo revolution*. Penguin UK.
Kracauer, S. (2014). *The Past's Threshold*, diaphanes
Lackey, Jennifer (2008). *Learning From Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.
Langton, R. (1993). Speech acts and unspeakable acts. *Philosophy & public affairs*, 293-330.
Lee, Timothy (2019) I created my own deepfake—it took two weeks and cost $552, *Ars Technica,* 16 December 2019**,** https://arstechnica.com/science/2019/12/how-i-created-a-deepfake-of-mark-zuckerberg-and-star-treks-data/
Lopes, Dominic (1996). *Understanding Pictures*. Oxford University Press.
Lopes, Dominic McIver (2015). *Four Arts of Photography*. Wiley.
MacKenzie, D., & Spinardi, G. (1995). Tacit knowledge, weapons design, and the uninvention of nuclear weapons. *American journal of sociology*, *101*(1), 44-99.
MacKinnon, Catharine. A. (1987). *Feminism unmodified: Discourses on life and law*. Harvard university press.
Makuch, Ben (2021) How Those Viral Tom Cruise Deepfakes Were Made, *Motherboard* 19th March 2021, https://www.vice.com/en/article/xgzb9d/how-those-viral-tom-cruise-deepfakes-were-made
Matthews, Taylor R. C. (forthcoming). Deepfakes, Intellectual Cynics, and the Cultivation of Digital Sensibility. *Royal Institute of Philosophy Supplement*.
McGlynn, Aidan (2016). Propaganda and the Authority of Pornography. *Theoria: Revista de Teoría, Historia y Fundamentos de la Ciencia* 31 (3):329-343.
McIntyre, L. (2018). *Post-truth.* MIT Press.
Meek, James (2019) Deepfakery, *London Review of Books,* 5 December 2019
Mills, C. (2007). White ignorance. *Race and epistemologies of ignorance*, *247*, 26-31.
Morris, Michael (2020). *Real Likenesses: Representation in Paintings, Photographs, and Novels*. Oxford University Press, Usa.
Mui, Tsing (2021) Lucasfilm Hires YouTube Deepfaker Who Fixed Its Weird-Looking Luke, Leia, and Tarkin Star Wars Cameos *The FPS Review,* 27 July 2021, https://www.thefpsreview.com/2021/07/27/lucasfilm-hires-youtube-deepfaker-who-fixed-its-weird-looking-luke-leia-and-tarkin-star-wars-cameos/
Nussbaum, Martha C. (1995). Objectification. *Philosophy and Public Affairs* 24 (4):249-291.

Öhman, Carl (2020). Introducing the pervert's dilemma: a contribution to the critique of Deepfake Pornography. *Ethics and Information Technology* 22 (2):133-140.
Ovadya, Aviv (2018) Fake news is about to get much worse. Here's a solution, TEDxmilehigh, https://www.ted.com/talks/aviv_ovadya_fake_news_is_about_to_get_much_worse_here_s_a_solution/transcript?language=en
Paris, B., & Donovan, J. (2019). *Deepfakes and cheap fakes.* Data and Society report https://datasociety.net/library/deepfakes-and-cheap-fakes/
Reuters (2022) Fact Check-Animation miscaptioned as if to show video of Ukrainian fighter jet shooting down Russian plane, *Reuters Fact Check* Febuary 22nd 2022, https://www.reuters.com/article/factcheck-animation-ukrainianjet-idUSL1N2V035G

Reuters (2020) Fact check: "Drunk" Nancy Pelosi video is manipulated, *Reuters Fact Check,* August 3rd 2020, ttps://www.reuters.com/article/uk-factcheck-nancypelosi-manipulated-idUSKCN24Z2BI
Rini, Regina (2020). Deepfakes and the Epistemic Backstop. *Philosophers' Imprint* 20 (24):1-16.
Rini, Regina & Cohen, Leah (forthcoming). Deepfakes, Deep Harms. *Journal of Ethics and Social Philosophy*.
Ritchin, Fred (2008) *After Photography,* WV Norton.
Roth, L. (2009). Looking at Shirley, the ultimate norm: Colour balance, image technologies, and cognitive equity. *Canadian Journal of Communication*, *34*(1).
Roth, L. (2019). Making Skin Visible Through Liberatory Design, in *Captivating Technologies,* , Ruha Benjamin (ed), Duke University Press, Durham: 275-307
Rothman, Joshua (2018) In the Age of AI is seeing still believing? *The New Yorker,* Nov 5, 2018 https://www.newyorker.com/magazine/2018/11/12/in-the-age-of-ai-is-seeing-still-believing
Schick, N. (2020). *Deep Fakes and the Infocalypse: What You Urgently Need to Know*. Hachette UK.
Shapin, S. (1989). The invisible technician. *American scientist*, *77*(6), 554-563
Shin, S. Y., & Lee, J. (2022). The Effect of Deepfake Video on News Credibility and Corrective Influence of Cost-Based Knowledge about Deepfakes. *Digital Journalism*, 1-21.
Sobande, Franscesca (2021). Spectacularized and branded digital (re) presentations of black people and blackness. *Television & New Media*, *22*(2), 131-146.
Sontag, S. (1977). *On photography*. Penguin, London
Steichen, Edward. (1903). "Ye Fakers." *Camera Work* 1 (Jan.): 448. Archived at https://modjourn.org/issue/bdr571938/
Tucher, A. (2017). "I believe in faking": The Dilemma of Photographic Realism at the Dawn of Photojournalism. *Photography and Culture*, *10*(3), 195-214.
Tucher, A. (2022). *Not Exactly Lying* Columbia University Press, New York.
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), 2056305120903408.
Wakefield, Jane (2022) Deepfake presidents used in Russia-Ukraine War, *BBC news* 18 March 2022 https://www.bbc.co.uk/news/technology-60780142
Walton, Kendall L. (1984). Transparent Pictures: On the Nature of Photographic Realism. *Critical Inquiry* 11 (2):246-277.

Warzel, Charlie (2018) Believable: The Terrifying Future Of Fake News, Feb 12 2018, *Buzzfeed News*, https://www.buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news