Catherine Kerner and Mathias Risse*

Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds

https://doi.org/10.1515/mopp-2020-0024
Published online November 12, 2020

Abstract: Deepfakes are a new form of synthetic media that broke upon the world in 2017. Bringing photoshopping to video, deepfakes replace people in existing videos with someone else's likeness. Currently most of their reach is limited to pornography, and they are also used to discredit people. However, deepfake technology has many epistemic promises and perils, which concern how we fare as knowers. Our goal is to help set an agenda around these matters, to make sure this technology can help realize epistemic rights and epistemic justice and unleash human creativity, rather than inflict epistemic wrongs of any sort. Our project is exploratory in nature, and we do not aim to offer conclusive answers at this early stage. There is a need to remain vigilant to make sure the downsides do not outweigh the upsides, and that will be a tall order.

Keywords: deepfakes, epistemic justice, epistemic rights, film, digital lifeworlds

1 The Brave New World of Synthetic Video

Suppose you hear Barack Obama call Donald Trump a "complete dipshit," Mark Zuckerberg boast about "control of billions of people's stolen data," or *Game-of-Thrones* protagonist Jon Snow apologize for the show's ending.[1] Chances are your source is a *deepfake*. Bringing photoshopping to video, deepfakes replace people

1 See, respectively, https://www.youtube.com/watch?v=cQ54GDm1eL0&feature=emb_logo; https://www.youtube.com/watch?v=Ox6L47Da0RY;https://www.youtube.com/watch?v=4GdWD0yxvqw&feature=emb_logo; last access May 2, 2020.

***Corresponding author: Mathias Risse**, Harvard Kennedy School, Harvard University, 79 JFK St, Cambridge, 02138, USA, E-mail: mathias_risse@harvard.edu
Catherine Kerner, Department of Computer Science, Harvard University, Cambridge, Massachusetts, USA

in existing videos with someone else's likeness. They are named after their usage of deep learning technology, a branch of machine learning that applies neural net simulation to massive data sets. Artificial intelligence learns what a source face looks like at different angles to transpose it onto a target, as if that target wore a mask. We propose a framework for thinking about some central epistemological and ethical issues that we ought to keep in mind so that humanity can enjoy the promises of the technology behind deepfakes ("deepfake technology") rather than suffer its perils.

While only time will reveal this technology's trajectory, we can identify some promises and perils to watch. These concern the way we acquire knowledge, and come to be known by others, in digital lifeworlds. "Lifeworld" (from the German *Lebenswelt,* which is familiar from especially Husserl's phenomenology) characterizes the impressions, activities and relationships that make up the world as a person experiences it and people in shared contexts experience it together.[2] Lifeworlds increasingly include, and are shaped by, electronic equipment and applications that use information in the form of code. Digital lifeworlds already connect humans, sophisticated machines and abundant data in elaborate ways. As science writer Jamie Susskind argues, digital lifeworlds are *pervasive* in that more and more devices do their tasks linked to the Internet; *connective* in letting people in far-flung locations interact; *sensitive* in that sensors trace ever more things and information; and *constitutive* in that machines are essential to our reality, rather than cyber add-ons to a life otherwise focused; and *immersive* by way of offering more and more augmented and virtual reality (Susskind 2018, chs. 1–2).

Digital lifeworlds enable new ways of acquiring knowledge, of shaping contexts in which acquisition happens, and of being known. They open up artistic possibilities unknown to the analog world. To all this change, synthetic media – media produced or modified through digital technology, especially artificial intelligence – will contribute enormously. Such media might personalize, and revolutionize, upbringing and personal development. For each learner amazing opportunities could arise through technologies that capture people, including oneself, in situations they never inhabited, or through recreating scenarios that so far could only be accessed through the use of faces other than those of the protagonists. We might soon conclude that – the downsides notwithstanding, which will definitely need appropriate regulation – "deepfakes" was an unfortunate choice of name, resonating primarily with associations with "fake news," which began to play its infamous role in American (and

2 We take the term "digital lifeworld" itself from Susskind 2018. For Husserl's work, see Smith 2013.

global) culture in earnest only with the 2016 presidential election campaign. Talking about "synthetic media" might be more conducive to getting the whole range of relevant issues into sight here.

So we need closer scrutiny of the *epistemic* promises and perils of deepfakes (i.e., those pertaining to the domain of *inquiry*) against the background of possibilities generated by digital lifeworlds. To set the stage, section 2 talks more about deepfakes and section 3 discusses some general epistemological issues around film. To make clear how scrutiny of epistemic promises and perils is always an inherently moral endeavor, section 4 introduces a framework of "epistemic actorhood" to capture different roles persons play in the exchange of information, with an eye on digital lifeworlds. First of all, people operate as *individual epistemic subjects*, knowers whose endeavors ought to respect certain standards of inquiry. Secondly, people are part of a *collective epistemic subject*, in which capacity they help establish or maintain such standards. Thirdly, persons are *individual epistemic objects*, getting to be known by others as delineated by rules concerning what information about oneself may be shared. Finally, individuals are part of a *collective epistemic object*, in which capacity they maintain or enhance the pool of what is known about us collectively and help decide what is done with it.

Using this framework, sections 5 and 6 introduce the notions of epistemic rights and epistemic justice, and with that vocabulary in place, sections 7 and 8 explore ways in which epistemic actors could be wronged in their various roles. But while there decidedly are such perils, deepfakes (or in any event the underlying technology) also come with some promises for each role. The range of both is substantial, though a lingering concern will be that there is simply not enough of an upside to this technology to make good on the downsides. At the very least, it will take much thought and careful regulation to make sure we can enjoy the promises without suffering too much damage from the perils, and to make sure that especially society's most vulnerable are protected from those perils. Also, media used to maintain epistemic actorhood (to bring about some kind of epistemic success) not only can be used to distort such actorhood (bring about epistemic failure), but also for non-epistemic, experimental purposes, like self-expression or self-discovery. Accordingly, section 9 explores creative uses of deepfake technology. Section 10 concludes. Our goal is decidedly not to come to bottom-line conclusions but to help set an agenda around reflecting on some epistemological and ethical issues we ought to keep in mind so humanity can enjoy the promises of an emerging technology. That agenda can only be executed fully as deepfake technology develops. On the technology side this paper reflects where things stand in early summer 2020. But the philosophical framework we propose should provide guidance for the debate as things unfold.

2 Deepfakes, Cheapfakes, and What All This Has to Do with Pamela Anderson

Deepfakes got started in 2017 – in 2020 the term is still recent enough for Word to underline it on the screen – when an eponymous Reddit user enlisted open-source software from Google and elsewhere to apply scattered academic research to face-swapping. They uploaded doctored clips mapping faces of celebrities such as Scarlett Johansson, Gal Gadot, or Taylor Swift onto bodies of porn actresses. Soon others in the Reddit community r/deepfakes shared creations of their own, with non-pornographic videos often having actor Nicolas Cage's face swapped into various movies.[3] Deepfakes came to public attention in December 2017, following an article in Motherboard by tech writer Samantha Cole.[4]

Discreditation is another area where deepfakes had an impact, as did the less sophisticated "cheapfakes," a coinage owed to media scholars Britt Paris and Joan Donovan. Doing without machine learning, cheapfakes are audio-visual manipulations created via Photoshop, the use of lookalikes, or the re-contextualization of footage or speeding up or slowing down of footage. Such efforts can make people appear incapacitated or as if they were moving faster or slower than they did to alter the nature of what occurred. In November 2018, CNN reporter Jim Acosta saw his credentials suspended after a cheapfake seemed to show him strike a White House intern when actually he had stayed her arm to hold on to a microphone in an exchange with Trump.[5] But while that video distorted a real event, Indian investigative journalist Rana Ayyub, in April 2018, found herself featured in a deepfake porn. Her face was swapped in, the actress, who was younger and had different hair. Still, the video went viral across India and created broad knowledge of "witnessing" Ayyub in an intimate setting or of "finding out" about a side-job in porn, undermining her standing as a journalist.[6]

---

3 https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them; https://www.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley; https://mashable.com/2018/01/31/nicolas-cage-face-swapping-deepfakes/; https://www.knowablemagazine.org/article/technology/2020/synthetic-media-real-trouble-deepfakes; last access May 2, 2020. For a good review of the technology and its emergence: https://timreview.ca/article/1282; last access May 24, 2020.

4 "AI-Assisted Fake Porn is Here and We're All Fucked." https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn; last access May 24, 2020.

5 https://datasociety.net/library/deepfakes-and-cheap-fakes/; https://slate.com/technology/2019/06/drunk-pelosi-deepfakes-cheapfakes-artificial-intelligence-disinformation.html;https://www.newyorker.com/news/current/the-white-houses-video-of-jim-acosta-shows-how-crude-political-manipulation-can-belast; last access May 2, 2020.

6 https://www.huffingtonpost.in/rana-ayyub/deepfake-porn_a_23595592/; last access May 10, 2020.

To be sure, researchers and special-effects studios have long pushed the boundaries of video manipulation. For instance, the 1994 film *Forrest Gump* (directed by Robert Zemeckis and starring Tom Hanks) used footage of JFK with altered mouth movements.[7] The story of video manipulability resembles that of photography: photos could be manipulated decades before digitalization and increasingly powerful software enabled any competent user to do as good a job as Stalin's specialists did editing out erstwhile allies after their fall from grace.[8] What Zemeckis and others did to video was expensive, time-consuming and required artistic skill. Soon, deepfake technology could enable anybody to make convincing videos featuring themselves or anybody else or pay companies that do the processing in the cloud rather than in a high-tech studio. Deepfake technology can also create photos from scratch to help create fictional online personas.[9] Audio can be deepfaked too to create "voice skins" or "voice clones" (digital assets that transform voices in real time, allowing anyone to speak as their chosen online persona).[10]

For now, non-consensual celebrity porn accounts for the lion's share of deepfakes, most others being jokes of the Nicolas Cage variety.[11] But recall the extraordinary role TV personality Pamela Anderson played in the spread of the Internet. Known through the 90s series *Home Improvement* and *Baywatch*, Anderson holds the record of most *Playboy* covers by any person. She was the most searched-for person on the Internet between 1995 and 2005. But her shows eventually became TV history, and though pornography appears to still swallow up 30% of Internet bandwidth,[12] the Internet outgrew the "original

7 https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/; last access May 2, 2020.

8 On this see King 2014. One case (which after multiple revisions only leaves Stalin) is in the public domain: https://commons.wikimedia.org/wiki/File:Soviet_censorship_with_Stalin2.jpg; last access May 24, 2020. It is worth noting that Winston Smith, the protagonist of George Orwell's dystopian novel *1984,* works as a clerk in the Records Department of the Ministry of Truth, rewriting documents to match the constantly changing party line. This involves revising articles and doctoring photographs, mostly to remove "unpersons," people who have fallen afoul of the party (see Orwell 1961).

9 A non-existent Bloomberg journalist, "Maisy Kinsley," who had a profile on LinkedIn and Twitter, was probably a deepfake. Another LinkedIn fake, "Katie Jones," claimed to work at the Center for Strategic and International Studies, but is thought to be a deepfake created for a foreign spy operation; see https://apnews.com/bc2f19097a4c4fffaa00de6770b8a60d; last access May 10, 2020.

10 https://modulate.ai/; https://slate.com/technology/2019/08/vocal-deepfakes-music-human-machine-collaboration.html; last access May 10, 2020.

11 https://deeptracelabs.com/mapping-the-deepfake-landscape/; last access May 10, 2020.

12 https://www.huffpost.com/entry/internet-porn-stats_n_3187682; last access May 13, 2020.

influencer."[13] It enabled new forms of activities and associations, ranging from networking and entertainment, electronic business, peer-to-peer philanthropy, telecommuting and collaborative publishing to politics and even revolutions. Similarly, in time the technology behind deepfakes is likely to have implications for our increasingly digital lifeworlds far beyond porn and discreditation (which should not belittle harms done in the meanwhile).

Deepfake detection in its current state is often referred to as a "cat-and-mouse" game, a term originally used to describe the competition between quickly evolving cybersecurity attacks and defenses. Here the adversarial game is between deepfake generators and the learned detectors designed to identify them. For example, one solution detects deepfakes on the basis of eye blinking, as deepfake generators rarely receive input frames with closed eyes. So subjects in deepfakes do not follow natural blinking patterns. But the researchers acknowledge that *the very publication* of their paper will likely ensure that serious forgers consider blinking from now on. Comments by the researchers who developed the eye-blinking detector make this adversarial mindset clear:

> Lyu says a skilled forger could get around his eye-blinking tool simply by collecting images that show a person blinking. But he adds that his team has developed an even more effective technique, but says he's keeping it secret for the moment. "I'd rather hold off at least for a little bit," Lyu says. "We have a little advantage over the forgers right now, and we want to keep that advantage."[14]

One naturally wonders when this escalation will reach its ceiling. Much of the forward-looking literature on deepfakes predicts the imminent arrival of this limit, the point in time when deepfakes obtain perfect photorealistic quality. Detectors, however perfect themselves, will then not be an effective solution. A report by The Brookings Institution calls that new state of affairs the cat-and-cat game.[15]

3 Capturing Reality: The Epistemology of Film

In 1896, Louis Lumière released one of the first motion pictures ever, *L'Arrivée d'un Train en Gare de la Ciotat*. Only 50 s long, the film captures an unremarkable scene:

---

13 https://www.hollywoodreporter.com/features/pamela-anderson-defends-julian-assange-talks-vladimir-putin-more-1107298; https://ftalphaville.ft.com/2019/05/10/1557479303000/Alphaville-meets-Pamela-Anderson–the-original-influencer/; last access May 2, 2020.

14 This is quoted here: https://www.technologyreview.com/s/611726/the-defense-department-has-produced-the-first-tools-for-catching-deepfakes/; last access May 24, 2020.

15 https://www.brookings.edu/research/fighting-deepfakes-when-detection-fails/; last access May 24, 2020.

a steam engine arrives at a station, passengers disembark, others board. With the camera set at a low angle to the tracks, the train grows larger and larger in the frame until it appears as if the locomotive might barrel into the theater. The movie entered the annals of film owing to accounts that audience members screamed or even fainted in the face of the onrushing train. Much of that account has by now been identified as the "founding myth of film" (Loiperdinger and Elzer 2004).[16] But like photography, film – originally a rapid sequence of photographs – has had an impact precisely because "it is so real." The epistemic value of photographs stems from their use as recordings, true accounts of how things are. Kendall Walton explained the epistemic value of traditional film photographs by likening cameras to mirrors. By reflecting light, mirrors enable us to see objects outside our line of sight, around a corner for example. Similarly, cameras capture light and enable viewers to see through time and across distances. Viewers can really "see" objects through photographs, if only indirectly (Walton 1984).[17]

Walton's "transparency thesis" – that photographs enable literal perception – grounds much philosophical work on film in the analytical tradition. To be sure, much skepticism resulted about how much three-dimensional objects at time t_1 could be like two-dimensional images at t_2. This has led to improved attempts at capturing the "realism" associated with photography. Dan Cavedon-Taylor, for one, plausibly argues that the advantage of photography over painting is that the former generates *perceptual* knowledge while the latter can only support *testimonial* knowledge. Testimony leaves more space for doubt than perception does. As Cavedon-Taylor put it, "the conditions under which it is rational to believe the content of another's testimony are stricter than those under which it is rational to believe the content of another's photograph" (Cavedon-Taylor 2013, pp. 288–89).

What is most interesting for our purposes is Walton's reasoning for his transparency thesis, which is a view about the technology behind film. The process of capturing and developing a traditional photograph (and films drawing on that process) is mechanical, so the experience of viewing it is connected causally to subjects in the real world. This causality puts viewers "in contact" with objects in photographs the same way they would be if they were viewing them in real life. Knowledge can be as reliably acquired through seeing something in photographs as it can by sight perception. Robert Hopkins recently revised that view to take into account doubts about the kind of encounter photography makes possible (Hopkins

16 The "Roundhay Garden Scene," recorded in 1888, is believed to be the oldest surviving film in existence; https://www.thevintagenews.com/2016/01/10/roundhay-garden-scene-is-believed-to-be-the-oldest-known-video-footage/; last access May 24, 2020.

17 Analytical philosophy came late to film. In continental thought more work has been done (starting already in the 1930s), drawing especially on Benjamin (2008).

2012). Photographs are epistemically valuable because they present us with putative facts, generating a "factive pictorial experience." That experience draws on causal processes of light capture and development used to produce film photographs, which, accordingly, *represent* objects in the real world. The facts they present us with cannot represent the world in ways other than it is or was. It is because photography is guaranteed to be factive that it is a reliable source of knowledge, of not only true but *justified* belief.

By way of contrast, digital photography, which accounts for almost all image media consumed today, is *capable* of being fact-preserving, but does not guarantee the factivity of traditional photography. Digital images are captured by an entirely different process, one Hopkins does not deem appropriately causal. He claims a subprocess called interpolation, an engineering shortcut behind digital image capture, makes synthetic media incapable of factive guarantees. Also, owing to how they are stored, digital photographs are more easily manipulated than film photographs, rendering manipulated specimens indistinguishable from unmanipulated ones. Hopkins worries that it is possible to create a digital image out of a set of pixels such that they match exactly what cameras would capture if the scene were real. That is what deepfakes do now. As Barbara Savedoff warned not long after consumer digital photography first appeared:

> If we reach the point where photographs are as commonly digitized and altered as not, our faith in the credibility of photography will inevitably, if slowly and painfully weaken, and one of the major differences in our conceptions of paintings and photographs could all but disappear. (Savedoff 2000, p. 202)

To be sure, certain epistemological limitations of film – in addition to the fact that, with much effort, they too could be forgeries – have long been known, though sensibly have not broadly undermined its authority. Let us mention two. To begin with, anthropologists around the turn of the twentieth century enthusiastically embraced film to study non-Western cultures. But they soon realized that film could not create a deep enough understanding of how people interacted in cultural contexts utterly discontinuous with the viewers' own. Whatever impact Lumière *L'Arrivée* had, it could have it only because viewers knew trains and train stations. Film can connect audiences with "what really happens," but only if they have a suitable context. Anthropologists soon switched their methodology to immersive fieldwork, producing monographs instead (Griffiths 2001, ch. 4).

Consider another issue. Abraham Zapruder's famous film of the assassination of JFK was used to check on the testimony of thousands of eyewitnesses, each of whom was sure of dramatically different things. Complexities of speed, emotion, distance, and memory made it hard to judge which testimony to trust. Zapruder's

film let investigators build a single narrative (Wrone 2003). However, they wrongly assumed the film captured the entire assassination. Instead, the first shot had been fired before Zapruder's camera was on. Trying to interpret all three shots within the film's timeframe led to inconsistencies that were seized upon by conspiracy theorists. The underlying problem is one of over-reliance on the epistemological virtues of film.[18]

4 Epistemic Actorhood

With these basic epistemological issues about film in view, let us proceed to our model of epistemic actorhood. For present purposes we understand information in terms of data. Data is anything recorded and transmissible in some act of communication. Information is data that is *useful* in given contexts. Most commonly (and minimally) data is useful by being accurate, the kind of thing captured in truthful statements.[19] Information-gathering with the intention of acting back on the environment is the key activity of intelligent life. For humans, inquiry – the systematic gathering of information through language or otherwise – is an essential pursuit. Historian David Christian calls us "networking creatures" to emphasize the extent to which *collective* learning characterizes our species (Christian 2004, part III). Collective learning grounds all culture. Once scripts are available, information can be preserved accurately and can grow over generations, amplifying our ability to shape our environment.

Much scrutiny is devoted to what constitutes successful inquiry, involving fields like epistemology or scientific methodology. However, knowledge acquisition arguably is not exhaustively understood as a purely rational matter. Inquiry inevitably occurs in contexts where information is channeled and presented in some manner, and where it is more or less difficult for people to acquire knowledge, including self-knowledge. Scrutiny of inquiry therefore also involves fields like intellectual history, ethics, sociology or political science. Throughout history and across cultures, multifarious standards of inquiry evolved. Michel Foucault used the term *episteme* – Greek for understanding – to denote the structure of thought, or the worldview(s), of an era, structures that, one way or another, are collectively maintained in ways that reflect power structures and that individual inquirers can evade only under great effort, intellectual or political. The episteme

18 See Holland 2014, https://www.newsweek.com/2014/11/28/truth-behind-jfks-assassination-285653.html; last access May 24, 2020. This topic is very helpfully discussed by Rini (2020).

19 On the fascinating histories of the notions of data and information, respectively, see Rosenberg 2013; Peters 1988.

includes a shared set of rules of how to conduct inquiry, and of who gets to conduct what kind of inquiry, as well as a shared body of what counts as knowledge.[20]

But as we reflect on inquiry, we must recognize humans not merely as individual knowers who collectively maintain epistemes, but also who (wittingly or unwittingly) reveal information (again both individually and collectively). Much of the information people seek is about other humans. So individuals – things about them, personal data – are *known to* others. Individuals are "knowers," but also "knowns." And people are also known in aggregates: individuals gather information about behavioral patterns of neighbors, customers or fellow citizens. Polling and market research have made strides in coming to know people collectively, for which digital lifeworlds offer plenty of tools. As revealers or bearers of information, individuals are subject to rules that define success in terms of known-ness, one's own and that of others. These rules are a subset of those that apply to successful inquiry (where then the target of inquiry is humans). What is distinctive about this subset is not the separate rationality that applies to seeking information, but the moral, social or political standards that apply to what information should or should not be available about people, and to whom. Moreover, as members of collectives, people maintain such rules of revealing and also the content of what is known about us (all of which, again, is part of the episteme, since knowers are also knowns).

An "epistemic actor" is a person or entity integrated into some communication network (some system of information exchange) as a seeker or revealer of information. In academic discourse, "actors" are normally people with agency ("agents"), connoted with terms like choice or rationality. But in ordinary parlance "actors" often are performers who follow scripts. This sense of "actor" is what we enlist. Talking about epistemic actors rather than agents deliberately de-emphasizes that they *do* things. Epistemic actors have thoughts, feelings and beliefs: they *are* certain ways that can become known. Moreover, in terms of what occurs within networks, seekers and revealers obtain or generate information according to prevalent standards, which vary in nature from rational to moral or sociological. These standards can be critically assessed or transgressed, but normally individuals – the actors – do not even make a noticeable contribution to them. They fill roles by meeting expectations.[21]

We can distinguish four roles that constitute epistemic actorhood: individual epistemic subjects, collective epistemic subjects, individual epistemic objects, and collective epistemic objects. Since we are interested in digital lifeworlds, we introduce these roles with an eye on such contexts. First of all, people operate as

20 See Foucault 1982, 1994, 1980. On Foucault, see Watkin 2018; Gutting 2001, ch. 9.

21 So we use "actorhood" in the sense in which it is used by sociologist John Meyer in his world-society approach, see Krücken and Drori 2010.

individual epistemic subjects: they are learners or knowers whose endeavors are expected to respect certain standards of inquiry, ranging from standards of rationality (how best to obtain information) to moral standards or plain societal divisions of labor (who is supposed to have what kind of knowledge). To gather and process information, people must figure out established norms within the episteme. This will include finding appropriate use for media, ranging from books or newspapers to photos or videos. In digital lifeworlds the role as such has been transformed since the way we gather information has been affected considerably through the availability of digital media. We may google things, or have information sent our way from certain platforms. Information is now stored and processed on an astronomic scale, and the Internet has started to approximate something like H. G. Wells' *world brain* (Wells 2016).

Secondly, people are part of a *collective epistemic subject*, in which capacity they help establish or (more commonly) maintain standards of inquiry, the various types of rules constitutive of the current episteme. Whereas in the first role we ourselves figure things out, according to certain standards, in this second role we hold others to certain standards and help create those. So this role is about the maintenance of the episteme. For many people the ways in which they fill the role of contributor to or sustainer of the information environment is rather passive, typically consisting in compliance.

Thirdly, persons are *individual epistemic objects*, getting to be known by others as delineated by rules concerning what information about oneself may be shared. This role is that of an information holder, or provider, the role of a *known*. It is about managing privacy, with its many complications. Expectations around the role of individual epistemic objects apply to oneself and to others: there are limits to what we may reveal about ourselves (which depend on whom we interact with), and there are expectations around what we may reveal about others, or otherwise ways in which we make it possible that they get to be known in certain ways. What we feel or believe itself increasingly is data that can be gathered or inferred from other things we do (such as clicks). We can be tracked and traced in all sorts of ways. We are subject to much surveillance.[22] Accordingly, this role has been much boosted through the transition to digital lifeworlds. People may even become celebrities through the way they open up about themselves (and thus become influencers).

Finally, individuals are part of a *collective epistemic object*, in which capacity they maintain and contribute to the pool of what is known about us collectively and help ascertain what is done with it. This last role is that of a contributor to data

22 For recent discussion, see Zuboff 2019. For the advice that a proper response to data surveillance is obfuscation, the deliberate addition of ambiguous, confusing or misleading information to interference with surveillance and data collection, see Brunton and Nissenbaum 2016.

patterns, parallel to that of the maintainer of the epistemic environment in which information is gathered. Digital lifeworlds have brought lasting changes to data-gathering because we can now be known collectively in ways that draw on an immense pool of indirectly inferred information about our inner lives and private acts that nonetheless gives rise to known patterns of human behavior, thought, and feeling. This kind of understanding of human patterns would have been previously unthinkable.

5 Epistemic Rights

With this framework in place we introduce some normative notions capturing ways in which epistemic actors are wronged in their roles. At least some of the wrongs people might experience as *individual* epistemic subjects and objects can be assessed in terms of violations of epistemic rights. By way of contrast, wrongs people might suffer as part of collective subjects or objects are structural failings, and thus often plausibly captured as violations of epistemic justice. Using this vocabulary allows us to formulate certain moral demands as they apply in the domain of inquiry.

Let us begin with epistemic rights and how they bear on the two individual roles. At a general level, rights are entitlements that justify the performance or prohibition of actions by the right-holder or another party. Philosophers have long distinguished among several types of rights, which may be components of the same right. In terms of the widely accepted Hohfeldian scheme, rights might be privileges, claims, powers or immunities (Hohfeld 1919; see also Wenar 2015). For there to be something sensibly called epistemic rights, there has to be a range of objects (broadly understood) of which individuals are aware and to which individuals may have differential entitlements; and which are of sufficient collective interest to merit efforts to limit access to them. Most straightforwardly, this kind of object would be *information*. Epistemic rights are rights that concern who is entitled to what kind of information.

We can illustrate this notion with regard to individual epistemic subjects and objects. As far as the former case is concerned, suppose I am tested for a disease. Normally I should be allowed to inquire about my results. That is, I have a *privilege-right* to know the result (no duty not to). I also have a *claim-right* against the provider to learn my result: they have a duty to inform me, rather than not to do so or to misinform me. And I have a *power-right* to waive my claim-right and thus not to know my result. Finally, I have an *immunity-right* that protects me from the provider altering my entitlements with regard to this information. There might be reasons entitlements should be regulated some other way, but the point here is to

illustrate how this notion of an epistemic right operates for individual epistemic subjects.[23]

Let us illustrate how this notion applies in the case of individual epistemic objects (the known). To continue with the test scenario, normally nobody else will have a privilege-right to know the results. Others have a duty to refrain from investigating the matter. It is my privilege-right not to be known to others in ways that include my results. Accordingly, nobody else normally has a claim-right against the provider to learn my results. I have a power-right to entitle other parties to know my result. Finally, an immunity-right protects me from any other party altering entitlements with regard to this information.

But while epistemic rights (concerning knowers and knowns) are most readily understood in terms of information, we may evoke the distinction among various *epistemic successes* (or *epistemic goods*) to substantiate talk of rights to know, to true and justified beliefs, to understand, or to truth (in the case of epistemic subjects), or of rights to privacy, to be forgotten, or rights against slander or theft of information. In all such cases one would need to spell out what type of right is meant (privilege, claim, power, immunity), and in what domain of data these rights operate. I might have a right to know my test results, and to that extent have the rights to understand my health situation, or the right to the truth in that regard. But I would have no right to the truth about other things, like other people's results. Similarly, I might have a right to privacy as far as my data are concerned, or the right that my data be deleted from certain places. But I might have no such rights as far as other matters are concerned, like the sale price of my home, which for good reason might be on public record.

Epistemic rights are confined to the domain of inquiry: beyond *learning of* something I might not be entitled to doing anything with it. Perhaps I am not even allowed to share it myself, or am in no position to market it in any way, etc. Similarly, beyond being entitled to have my information protected in certain ways I might have no further claims against people I interact with. Epistemic rights are *sui generis* and not naturally reducible to any other type of right (such as property rights).

Epistemic rights justify performance or prohibition of certain actions in the domain of epistemic goods, and we can apply them to both individual epistemic subjects and individual epistemic objects. But people can also be wronged in ways that involve structural features of communications networks. They would then

23 In this account of epistemic rights we follow Watson (2018). Earlier philosophical discussion about such rights explored the nature of epistemic justification, e.g., Dretske 2000. The point was to assess what kind of statements somebody is entitled to make even if they cannot do the work to justify them. See also Wenar 2003.

presumably be wronged as members of collective epistemic subjects or objects, rather than (exclusively) as individuals whose entitlements are thwarted. In such cases the language of justice would be appropriate, rather than that of rights.

6 Epistemic Justice

Let us first explain how talk about epistemic justice relates to a generic understanding of justice and other, better-known kinds of justice. The perennial quest for justice is about making sure each individual has an appropriate place in what our uniquely human capacities permit us to build, produce, and maintain, and that each individual is respected appropriately for their capacities to hold such a place to begin with. Under this umbrella we can distinguish *commutative* from *distributive* justice. The former maintains and restores an earlier status quo that set the stage for the interaction or otherwise responds to violations. The latter is concerned with sharing out whatever a community holds in common. Major themes in the history of reflection on distributive justice have been to assess just what the community holds in common, and what the relevant community is to begin with. An influential contemporary proposal, owed to Rawls, is to see the state as that community, and what that community holds in common are social primary goods: rights and liberties, opportunities and powers, income and wealth, and the social bases of self-respect.[24]

A notion of *epistemic* justice can be readily integrated, in light of the crucial role information plays in everything humans build, produce, and maintain. Epistemic justice is part of distributive justice, concerned with access to, and with making sure each individual has an appropriate place regarding, information. To be sure, the term "epistemic injustice" was introduced by Miranda Fricker to identify *wrongs* to people in their capacity as knowers, rather than embedding it into a larger understanding of *justice* (Fricker 2009). But to be clear about the distinction between epistemic rights and epistemic justice we need to know how epistemic justice relates to the broader context of justice-related discourse.

One might worry that capturing access to information in terms of justice is peculiar because much information is private and should be more sensibly discussed, say, under a heading of personal integrity. However, the connection to justice becomes clearer once we recall Foucault's insight that an episteme also includes self-knowledge. People's sense of self (and private information) is *constituted* by what social relations make possible, and thus falls under the nexus

24 For distributive justice, see Risse 2020. For this proposal, see Rawls 1971, 2001. For cyberspace as a site of justice, see also Duff 2013.

between power and knowledge that drives Foucault's thinking. Power structures determine who has access to established knowledge (including self-knowledge), but also affect what questions get asked, who is entrusted with exploring them, etc. One challenge then is to identify the nature of wrongs occurring at this nexus between knowledge and power. Digital lifeworlds are replete with possibilities of inflicting such wrongs.[25]

Epistemic injustices – wrongs around access to, and thus around individuals having an appropriate place regarding, information – can assume multiple forms. Let us present some examples in terms of how they apply to either the role of collective epistemic subjects or that of collective epistemic objects. A first example is when a group lacks adequate *access* to education, which typically would be women, minorities, or people at the lower end of the economic ladder. One may see such exclusion as lots of violations of epistemic rights. But that move misses a structural concern: as collective epistemic subject, we are systematically limiting access to information for certain groups, for the sake of maintaining power relations. Members of the excluded groups will often be unable to acquire requisite skills to be political and economic participants, which normally is the intention of the exclusion. Specifically for digital lifeworlds lack of education will normally entail a highly diminished capacity to participate, in anything other than a mostly passive role. So the more our lifeworlds turn digital, the graver an injustice denial of education is.

A second example is *testimonial injustice*, where certain speakers have diminished credibility because recipients have prejudices about their background or social group.[26] A narrow understanding is disregard of testimony to a court, where, say, due to racial prejudices an all-white jury might refuse to believe black witnesses. But given how much orientation in the world we acquire through testimony (broadly conceived) – most of what we think we pick up from our social context – testimonial injustice also occurs if perspectives are dismissed in day-to-day exchanges, are invisible in textbooks or side-lined in memory culture. What is peculiar about digital lifeworlds is that occasions for inflicting such injustices directly are increasingly avoided through the on-line echo chambers that arise if people decide what information they want delivered. Accordingly, the typical form of digital testimonial injustice is that these choices themselves reflect and reinforce

25 For the connection between Foucault's approach and information, see Koopman 2019. For the centrality of information for understanding human abilities, see Tegmark 2017, chapters 1–2. For the view that the ubiquity of global communication flows in the present age has collapsed the separated space needed for critical reflection (and thus in particular undermined anything that might credibly be critical theory), see Lash 2002.

26 On testimony, see Lackey 2010; Coady 1992; Goldberg 2010.

all prejudices society has nourished over time while creating relatively few situations where acts of injustice are committed to people's face. So even in principle there is no opportunity to examine the underlying prejudice in the presence of all concerned. The collective epistemic subject of digital lifeworlds is increasingly fragmented. The collective epistemic object is of a sort where people are known only through lenses of fragmented processing.

A third example is *silencing*. Silencing is the removal of one's ability to communicate through the creation of conditions under which one's utterances or speech acts are disregarded. The term came into circulation through discussions about how pornography objectifies women in ways that imply they are "not heard" when refusing sex (Langton 1993; Langton and Hornsby 1998; MacKinnon 1987). Silencing also extends to politics when outlandish claims are made about public figures to such an extent that we would have no reason to believe what they say in any situation (*mutatis mutandis* for other domains). To this form of epistemic injustice digital media provide new outlets, for instance through competition for the wittiest short statement of a view that has been cultivated through the prominence of Twitter in public life. Often no amount of reasoned speech can offset a cleverly worded two-line dismissal. And again this is a problem for how we acquire knowledge and for how we are known in the world.

And a fourth example worth mentioning is race-/nation-/gender-driven *ignorance*, the impact of class, gender, race etc. on belief acquisition. What is meant is the formation of mistaken beliefs because of the suppression of pertinent knowledge within certain populations. This phenomenon might arise even without prejudicial attitudes. A prominent example that has recently received much attention is *White Ignorance*. White Ignorance occurs if absence of pertinent knowledge among white people about the historical trajectory of people of color (especially in countries with a fairly recent history of slavery, like the US) precludes white people from comprehending the extent to which people of color saddled with disadvantaged starting points in life (Mills 2017, 2007).

In the domain of Big Data, there has been much discussion of this phenomenon. To begin with, those who work in IT are disproportionately from certain segments of society and ask questions about data that reflect their experiences. Secondly, data collection might occur through devices that certain segments of the population own more commonly than the population as a whole. Thirdly, the data themselves reflect what are often racist trajectories. In such ways the prejudicial structures of the past might end up shaping the future. And discrimination might be much harder to recognize if it is driven by factors correlated with odious phenomena, rather than by those phenomena directly (Barocas and Selbst 2016; see also Benjamin 2019).

7 Deepfakes and Epistemic Wrongs: Individual and Collective Epistemic Subjects

With these various ways of inflicting wrongs in place, let us see how deepfakes might inflict wrongs in terms of the four roles of epistemic actorhood distinguished earlier. Our main point is that for each role distinctive wrongs are created that we can capture in terms of violations of epistemic rights or epistemic injustices. But each time, there are distinctive gains we can also capture in terms of the realization of rights and justice. The challenge is to minimize the harms and cultivate the benefits, though there will be a lingering doubt that this can be done.

Let us begin with individual epistemic subjects. Inquirers are wronged if they have epistemic rights to particular information but receive deepfakes that provide false or misleading information. Straightforward examples are videos that misrepresent how events unfolded, for instance the Russian attacks on Syria.[27] In addition, to the extent that deepfakes become more widespread, individual epistemic subjects are not merely wronged in *particular* instances when they fail to receive information they have a right to. They are also wronged in their *overall role* as knowers to the extent that their ability to perform any of the tasks for which they need to be knowledgeable declines. Inquiry simply becomes harder to complete with more parties aiming to undermine it.

But deepfakes can also empower people as knowers and make it easier for them to realize epistemic rights. Consider three examples. To begin with, deepfakes can stipulate interest in fields like art and history by making them come alive. For instance, the Dalí Museum in St. Petersburg, Florida, used deepfake technology as part of an exhibition called Dalí Lives. To make good on that title, the museum created a life-size deepfake of the artist via a thousand hours of machine learning of his interviews.[28] This recreation could deliver a variety of statements Dalí had spoken or written. To mention another example, the Scottish company CereProc trained a deepfake algorithm on recordings of JFK. They could thereby produce a delivery of the speech he was due to give the day he was

27 See https://www.theatlantic.com/international/archive/2018/04/russia-syria-fake-news/557660/; https://www.reuters.com/article/us-mideast-crisis-syria-ghouta-provocati/russia-says-britain-helped-fake-syria-chemical-attack-idUSKBN1HK24P; last access May 24, 2020. For Russian information politics, see also Snyder 2019, ch. 5.

28 https://www.theverge.com/2019/5/10/18540953/salvador-dali-lives-deepfake-museum; https://www.forbes.com/sites/simonchandler/2020/03/09/why-deepfakes-are-a-net-positive-for-humanity/#442b6a2f2f84; last access May 10, 2020.

assassinated.[29] Secondly, deepfakes might convey messages more effectively. In 2019, a British health charity used deepfake technology to have David Beckham deliver an anti-malaria message in nine languages. Global celebrity might be dispatched effectively to convey information.[30] Thirdly, voice-cloning deepfakes can restore voices when people lose them to disease.[31] So in all these ways inquiry, and thus the exercise of epistemic rights, becomes easier through deepfakes. (But a sensible reaction to such examples would be that they fall far short of making good on the harms done).

As far as the collective epistemic subject is concerned, the main impact is the changing role of video in providing testimony. To begin with, deepfakes could allow people to produce recordings of events that never occurred, putting the burden on courts or competing parties to disprove that evidence. That could affect everything from custody battles or employment tribunals to criminal cases in which fake videos could provide alibis. Deepfakes could also mimic biometric data, tricking systems that rely on face, voice, or gait recognition. Similarly, deepfakes could be presented as long-lost evidence for an untenable position that some are nonetheless eager to believe. For example, some people question facts around the Holocaust, the moon landing and 9/11, despite available video proof of these events. Deepfakes could spread "alternative" versions, presented as long-suppressed evidence.

Moreover, the sheer possibility of deepfakes would create a plausible deniability of anything reported or recorded. Thereby doubts sown by deepfakes could permanently alter our trust in audio and video. For instance, in 2018, Cameroon's minister of communication dismissed as fake a video Amnesty International thinks shows Cameroonian soldiers executing civilians.[32] Similarly, Donald Trump, who in a recorded conversation boasted about grabbing women's genitals, later claimed the tape was fake. He thereby enabled his followers to take that stance.[33] Such denials would then be among the multifarious voices on an issue, making it ever harder to motivate people to scrutinize their beliefs.

In recent decades video has played a distinguished role in human inquiry, specifically in the context of testimony (in the broader sense discussed before).

---

29 https://www.bbc.com/news/uk-scotland-edinburgh-east-fife-43429554; last access May 10, 2020.

30 https://abcnews.go.com/International/david-beckham-speaks-languages-campaign-end-malaria/story?id=62270227; last access May 10, 2020.

31 https://www.projectrevoice.org/; last access May 10, 2020.

32 https://www.amnesty.org/en/latest/news/2018/07/cameroon-credible-evidence-that-army-personnel-responsible-for-shocking-extrajudicial-executions-caught-on-video/; last access May 13, 2020.

33 https://www.theguardian.com/us-news/2017/nov/29/denying-accuracy-of-access-hollywood-tape-would-be-trumps-biggest-lie; last access May 10, 2020.

What was captured on film served as indisputable (or least disputable) evidence of something in ways photography no longer could after manipulation techniques became widely available. Until the arrival of deepfakes videos were trusted media: they offered an "epistemic backstop" in conversations around otherwise contested testimony, as Regina Rini put it (2020). Without such a backstop, it will be hard to maintain trust that comes from reliance on established facts. Alongside other synthetic media and fake news, deepfakes might help create a no-trust society, in which people cannot, or no longer bother to, separate truth from falsehood, and no reliable media could help them do so. Within a few generations, people might no longer approach disagreements with a possibility of truth-finding in mind. This would then also be a society where the varieties of epistemic injustice, in their application to digital lifeworlds, would be pronounced, especially testimonial injustice.

How problematic is this, all things considered? Consider a related scenario. In their intricate discussion of "alphabetization" – the penetration of human culture by the written word, the advent of literacy – Ivan Illich and Barry Sanders discuss the changing role of the oath. "In the realm of orality," they explain,

> one cannot dip twice into the same wave, and therefore the lie is a stranger. My word always travels alongside yours; I stand for my word, and I swear by it. My oath is my truth until way into the 12th century: The oath puts an end to any case against a freeman. Only in the 13th century does Continental canon law make the judge into a reader of the accused man's conscience, an inquisitor into truth, and torture the means by which the confession of truth is extracted from the accused. Truth ceases to be displayed in surface action and is now perceived as the outward expression of inner meaning accessible only to the self. (Sanders and Illich 1989, p. 85)

What they expound is how the oath ceased to be the epistemic backstop it could be in a world of orality. To be sure, to this day, the oath – and, for that matter, the signed statement – has special legal importance, not as an epistemic backstop but as a way of incurring special legal responsibilities. If indeed Illich and Sanders are correct – whether they are we cannot judge – there might well not have been any such (broadly accepted) backstop between the demise of the oath in that function in the twelfth century and the advent of photography in the nineteenth century.

Accordingly, we cannot take for granted that there is a backstop in investigations of testimony to begin with. In times where there is not, judgments have to be made relying on the track record of, and one's willingness to trust, the source of the testimony, or else one would have to undertake a thorough investigation of many background factors (witnesses, corroborating evidence, consistency with things known, etc.). In some ways our testimonial practices might revert to such a world through the perfection of synthetic video. The difference is that in

the earlier scenario we simply have no indisputable media that connect us to reality; in the future we have them, but their results can also be fabricated synthetically. One way of seeing how much of a loss that would be is to consider that there was an epistemic backstop for that part of history when democracies in territorial states became a widespread model at the global scale. In some ways it is reassuring that by historical standards such a backstop was not normally part of the episteme: our ancestors could manage without it. But they did not have to navigate the intricacies of large territorial democracies given the complexities a technological age makes possible.

But with all that said, deepfake technology also has upsides for the collective epistemic subject, the ways in which collectively we acquire knowledge. Deep generative models raise new possibilities in medicine and healthcare, such as the use of deep learning to synthesize data that will help researchers develop new ways of treating diseases without using actual patient data. "Fake" MRI scans have already been created. By training on these medical images and on 10% real images, these algorithms became as good at spotting brain tumors as algorithms trained only on real images.[34] In the medical world synthetic data could also help immensely with anonymization. It is often possible to identify individuals in anonymized datasets if ancillary sets can be cross-referenced. Synthetic data block such possibilities by "creating" new people.[35] New ways of generating knowledge thereby become available that enrich our episteme. Deepfakes as we know them would only be part of ongoing technological innovation, much as photos of Pamela Anderson were part of the spread of the Internet. At the same time, these benefits seem to come with much uncertainty and as of now feel a bit remote, whereas the potential damage to democracy is already foreshadowed in the way media are used and handled as of 2020.

8 Deepfakes and Epistemic Wrongs: Individual and Collective Epistemic Objects

As far as the role of individual epistemic objects is concerned, people are wronged in their capacity as knowns in the first instance through efforts to spread falsities about them. Their epistemic rights are violated: what is spreading about them is

34 https://www.ucl.ac.uk/news/2019/nov/opinion-how-technology-behind-deepfakes-can-benefit-all-society; https://www.fastcompany.com/90240746/deepfakes-for-good-why-researchers-are-using-ai-for-synthetic-health-data; last access May 10, 2020.

35 https://www.techworld.com/data/what-is-synthetic-data-how-can-it-help-protect-privacy-3703127/; last access May 10, 2020.

not how they should be known. But parallel to the individual epistemic-subject scenario, there is more. In that case the wronging occurred also through the creation of an environment where people can no longer operate as knowers. Similarly, in the case of individual epistemic objects, there is a rights-violation not only if actual falsehoods about that person (the object) are conveyed, but if the way in which anything pertaining to that person is conveyed undermines her ability to come to be known in appropriate ways.

Recall Rana Ayyub. Most people could detect that the woman in the video was not her. But the creation of a collective sense that now lots of people were "in" on something at least close to watching her in a sex act undermined her ability to come to be known the right way. Her dignity as a person was violated, her authority as a journalist undermined. This is the threat of deepfake porn: that women's fragile emancipation from being seen as sex objects more than as occupants of roles of professionals, citizens, or as human beings worthy of respect is undermined through depictions associated with objectification. Revenge porn has this effect, and normally on women more than on men because men are not emerging from this kind of role. Such a fate could await many women since now unsophisticated perpetrators would no longer require nude photos or sex tapes to threaten victims. They can manufacture them. Similarly, deepfakes could do damage to how people who come from groups that are still overcoming prejudicial history get to be known.[36]

Legally this will generally be hard to address. As *Wired* noted, "You can't sue someone for exposing the intimate details of your life when it's not your life they're exposing."[37] In deepfake porn, it would not be *that person's* body, and the face could be ever so slightly altered: everybody still realizes who it is but there is plausible deniability, just as with people who look naturally similar. However, all this might also change as we go forward and the possibility of attacks like the one on Ayyub becomes commonplace. Perhaps to some extent what happened to her was so effective because it was new, allowing lots of men to fold this abuse into fantasies of their own. If something can be done to everybody, seeing it done to one person might lose its thrill, and it would then be done less.

As collective epistemic objects, the way people generally come to be known changes. We all understand that we enter people's imaginations any number of ways. We come to be known to others in light of their prejudices, but also in ways that connect to their fantasies, traumas or dreams. But all along, these have been mental activities trapped in their minds unless they captured them in words,

36 For recent reflections on the relevance of reputation and the ways in which it comes about and is shattered, also with special attention to the Internet, see Origgi 2017.

37 https://www.wired.com/story/face-swap-porn-legal-limbo/; last access May 10, 2020.

drawings or paintings. Now we all come to be known to others conscious of the fact that we could enter into their artistic, possibly erotic, fabrications. ("We're all fucked," in terms of Samantha Cole's pathbreaking article on deepfakes.)[38] We are potential actors in somebody else's productions, though this will affect some persons – those who have ways of catching people's imagination – more than others.

Manipulated videos will also do damage to democracy: people are harmed as knowers, but also as knowns. Our general infrastructure of how we get to know people is changing, and is suffering, especially in a fast-moving political process where any fake news will take time to be rebutted. In the process, the various types of epistemic injustice can be readily inflicted.

But as far as the role of the collective epistemic object is concerned, there is also empowerment, much as there was in the case of the subject. Deepfake technologies can amplify things for which, appropriately, people should be known. For instance, during the 2020 Delhi Legislative Assembly election, the Delhi Bharatiya Janata Party used deepfake technology to distribute a version of an English-language advertisement by its leader, Manoj Tiwari, translated into Haryanvi (a Western Hindi dialect) to target voters from Haryana state, where that dialect is spoken. A voiceover was provided by an actor, and video of Tiwari's speeches was used to lip-sync the video to the voiceover.[39] Similarly, deepfake technology enables people to wear virtual masks on outlets like Snapchat to share experiences of abuse without revealing their identities. They could remain anonymous while retaining human features and the ability to convey emotion, thus preserving the essential humanity of survivors of abuse.[40] But once again, one might well leave this discussion with a lingering sense that the potential for more damage is enormous and already rather concrete whereas the benefits are much more uncertain.

9 The Creative Potential of Deepfakes

Epistemic actorhood is concerned with both the acquisition of knowledge and ways of being an object of knowledge. In these roles there can be success or failure

38 https://www.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn; last access May 24, 2020.

39 https://www.hindustantimes.com/india-news/bjp-s-deepfake-videos-trigger-new-worry-over-ai-use-in-political-campaigns/story-6WPlFtMAOaepkwdybm8b1O.html; https://www.financialexpress.com/india-news/bjp-used-deepfake-videos-of-manoj-tiwari-to-reach-out-to-voters-during-delhi-assembly-elections-report/1872651/; last access May 13, 2020.

40 https://www.vox.com/recode/2020/6/29/21303588/deepfakes-anonymous-artificial-intelligence-welcome-to-chechnya; last access July 7, 2020.

as far as inquiry is concerned. Inquiry occurs by means of certain tools, such as oral or written communication, imagery or video. Such tools can also be used for purposes that are not knowledge-related at all but are exploratory or artistic, concerned with self-expression or experimentation.

Language can capture accurate information (success of inquiry), and also, as a flipside, convey inaccurate or misleading information (failure of inquiry). But language can also tell stories, entertain or convey lessons about life, or be used in pursuit of the narrator's love for developing certain themes or linguistic playfulness. Similarly, images can capture or falsify reality, but also play with reality, or capture an author's imagination or sentiments about being in the world, without any intention to misrepresent anything and without anybody engaging with the image as a successful or failed attempt to capture reality.

Creative use of language or imagery not only allows people to escape into fictional worlds, it also helps constrain power in ways even the most relentless pursuit of truth never could, without falsifying anything. There is parody, satire and caricature, which have ways of furthering the realization of political equality by taking a humorous look at the powerful, perhaps ridiculing them to break through the seriousness that shapes power relations. Techniques that would be cruel when applied to the vulnerable or even to peers are liberating when applied to the powerful. As far as the word "parody" is concerned, its Greek origins are *para,* "beside, against," and *oide,* "song." Thus, the Greek word *parodia* has been taken to mean "counter-song," an imitation set against some original, presumably a song of praise of those already well-known.

Creative people have already discovered the potential of deepfakes, like German artist Mario Klingemann, a pioneer in the use of computer learning in art, known for work involving neural networks, code and algorithms.[41] But this potential goes much further. Anyone could have their likeness inserted into most any scenario available on the Internet or have somebody else's likeness inserted. This could involve sexual fantasies. But as the technology develops, much as these things unfolded in the development of the Internet (recall Pamela Anderson), this would be one among multifarious uses. That is, synthetic video applications would enable users to produce porn clips, and it might be difficult to prevent such apps from fulfilling this function. But they could also do any number of other things.

People fantasize about many things. They can capture their fantasies by using deepfake technology, or develop fantasies in videos from scratch. Many of the mind's wanderings could find new outlets. So far, visual storytelling is an expensive business. Hollywood studios spend billions on creating spectacles that

41 http://quasimondo.com/last access May 11, 2020.

transport audiences to other worlds.[42] Deepfake technology incorporates the ability to synthesize imagery, giving smaller-scale creators similar capacities for bringing imaginative creativity to life.[43] The common person's dream of a creative empire might materialize.

There is a thin boundary between inflicting an epistemic wrong by casting somebody in, say, a pornographic video produced and spread with the intention or net effect of undermining how somebody else is perceived, and the living-out of fantasies that would be part and parcel of an expansion of creative possibilities by deepfake technology. Legal regulation needs to draw the line. Much will depend on whether one's creation is spread. In that regard, deepfakes are not very different from how we generically think about fantasies in somebody's mind and their execution, which is mediated through a decision. Fantasizing and daydreaming should not be punishable, and are not offensive, even if dreamers avail themselves of deepfakes to capture their imagination. What we do *not* want is for such products to spread if such spreading could have pernicious effects on how somebody gets to be known. Also, virtual worlds have been around for a long time, and deepfake technology will give a big push to them and create new possibilities of connecting to people in distant places. The great advantage of the Internet has always been that it allows people in far-flung places to do things together. This technology will enhance that possibility.

More generally, deepfake technology might make it possible for us to live in a world where what people dream exists not merely in their minds but also in the cloud. It would be an enormous change in how people's inner lives relate to the outer world, in the sense that there is the possibility of extending one's mind in such ways without otherwise acting back on the environment in any way. We could then do things via cloud computing that so far we could only do in our minds or through paintings. The creative process as a whole thereby grows substantially.

The possibilities are immense, and worth exploring. Think about the TV series *The Crown* (about Elizabeth II) with faces of actual royals mapped onto faces of actresses and actors, or *Thirteen Days* with the real faces of the protagonists of the Cuban Missile Crisis. Actors would still be important, but not for standing in for historical figures of whom we have enough images to let them literally speak for themselves. The movie industry could not only improve dubbing on foreign-language films, but, more controversially, resurrect dead actors. James Dean is

42 https://www.globenewswire.com/news-release/2018/09/27/1577156/0/en/Global-VFX-Market-Will-Reach-USD-19-985-64-Million-By-2024-Zion-Market-Research.html; last access May 11, 2020.

43 https://techcrunch.com/2019/07/04/an-optimistic-view-of-deepfakes/; last access May 11, 2020.

already due to star in *Finding Jack*, a Vietnam war movie.[44] Some may wish for Clint Eastwood or a member of the Douglas clan to keep acting, and they may wish for the same. Some actors and actresses have become timeless, and with deepfake technology, it might appeal to many to continue to see them featured in movies they (presumably) would have wanted to be in. And if you wanted your movie to be narrated by Ronald Reagan, Morgan Freeman or Michelle Obama, you might just make that happen. Also, finally, the ability to mimic faces, voices, and emotional expressions is one of the most important steps toward building a believable virtual human we can actually interact with. That would come with a whole new set of possibilities.

10 Conclusion: Where Do We Stand?

Deepfakes are mixed news, with the negative aspects already more clearly in sight than any possible benefits. They bring change that will have positive and negative consequences as far as the various epistemic roles are concerned. Much thought and regulation will be required to make sure epistemic roles are strengthened rather than weakened, that epistemic rights and justice as well as human creativity are enhanced, rather than wrongs inflicted. And such regulation would especially have to make sure that society's most vulnerable receive special protection. At the macrolevel there is a risk of enormous danger to democracy, and at the microlevel whose dignity and standing could be undermined through deepfake technology. Accordingly, a lingering sense will remain that the promises ultimately cannot in any sense outweigh or even match the perils.

Deepfakes are recognized as dangerous even by those who build them. Companies like Reddit, Facebook or Twitter have adjusted policies in response to deepfakes. There has also been some notable legal action. An early mover, California passed two bills related to deepfakes in 2019. Assembly Bill 730 addresses the threat of deepfakes in influencing elections. It prohibits "distributing with actual malice materially deceptive audio or visual media" of a candidate 60 days before an election, unless it is indicated that the material is manipulated." This bill expires in 2023.[45] Assembly Bill 602 addresses pornographic forgeries. This bill does not sunset. It gives victims cause for action against anyone who

44 https://www.hollywoodreporter.com/news/afm-james-dean-reborn-cgi-vietnam-war-action-drama-1252703; https://www.thewrap.com/james-dean-to-be-digitally-reanimated-in-cgi-for-vietnam-war-movie-finding-jack/; last access May 14, 2020.

45 https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB730; last access May 24, 2020.

"intentionally discloses sexually explicit material that the person did not create if the person knows the depicted individual did not consent to its creation."[46] To be sure, California laws do not apply elsewhere and are unlikely to bear, say, on international election tampering. Virginia and Texas have also passed legislation.[47]

The Deepfakes Accountability Act is a proposed federal effort. Its primary effect would be to instate transparency requirements, demanding clear disclosures on manipulated content. As of March 2020, the bill has been introduced to the House, awaiting further action. Well intentioned, this bill is drafted by legislators whose plans to watermark all "advanced technological false personation record [s]" may be hard to implement.[48] The rift between legal action and real-world action was recognized in a House Intelligence Committee meeting in June 2019, which emphasized that the responsibility to label manipulated video should rest with platforms. Regardless of responsibility, it is likely that platforms have the expertise and ability to directly address video forgeries.

This is merely where things stand as of early summer 2020. Deepfake technology has come to stay, and raises a host of questions, some of them philosophical. Our goal has been exploratory, to help set an agenda around such questions, rather than offer conclusive answers at this early stage. But we will need to remain vigilant to make sure that the downsides do not outweigh the upsides, and that will be a tall order.

Acknowledgments: We are grateful to the political theory colloquium at the University of Hamburg and to a convening of the human rights and technology fellows at the Carr Center for helpful Zoom discussion of this material in June/July 2020.

References

Barocas, S., and A. D. Selbst. 2016. "Big Data's Disparate Impact." *California Law Review* 104 (3): 671–732. http://dx.doi.org/10.15779/Z38BG31.

Benjamin, R. 2019. *Race after Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity.

46 https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB602; last access May 24, 2020/.

47 https://capitol.texas.gov/tlodocs/86R/billtext/html/SB00751S.htm; https://lis.virginia.gov/cgi-bin/legp604.exe?191+ful+HB2678S1; last access May 24, 2020.

48 https://capitol.texas.gov/tlodocs/86R/billtext/html/SB00751S.htm; last access May 24, 2020.

Benjamin, W. 2008. *The Work of Art in the Age of Its Technological Reproducibility, and Other Writings on Media*, edited by M. W. Jennings, B. Doherty, and T. Y. Levin. Cambridge, Mass: Belknap Press: An Imprint of Harvard University Press.

Brunton, F., and H. Nissenbaum. 2016. *Obfuscation: A User's Guide for Privacy and Protest*. Cambridge, Massachusetts; London: The MIT Press.

Cavedon-Taylor, D. 2013. "Photographically Based Knowledge." *Episteme* 10 (3): 283–97. https://doi.org/10.1017/epi.2013.21.

Christian, D. 2004. *Maps of Time: An Introduction to Big History*. Berkeley: University of California Press.

Coady, C. A. J. 1992. *Testimony: A Philosophical Study*. Oxford, New York: Clarendon Press.

Dretske, F. 2000. "Entitlement: Epistemic Rights without Epistemic Duties?" *Philosophy and Phenomenological Research* LX (3): 591–606. http://dx.doi.org/10.2307/2653817.

Duff, A. S. 2013. *A Normative Theory of the Information Society*. New York: Routledge.

Foucault, M. 1980. *Power/Knowledge: Selected Interviews and Other Writings, 1972–1977*, edited by C. Gordon. New York: Vintage.

Foucault, M. 1982. *The Archaeology of Knowledge: And the Discourse on Language*. New York: Vintage.

Foucault, M. 1994. *The Order of Things: An Archaeology of the Human Sciences*. New York: Vintage.

Fricker, M. 2009. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford, USA: Oxford University Press.

Goldberg, S. C. 2010. *Relying on Others: An Essay in Epistemology*. Oxford: Oxford University Press.

Griffiths, A. 2001. *Wondrous Difference*. New York: Columbia University Press.

Gutting, G. 2001. *French Philosophy in the Twentieth Century*. Cambridge, U.K.; New York: Cambridge University Press.

Hohfeld, W. 1919. *Fundamental Legal Conceptions*, edited by W. Cook. New Haven: Yale University Press.

Hopkins, R. 2012. "Factive Pictorial Experience: What's Special about Photographs?" *Nous* 46 (4): 709–31. https://doi.org/10.1111/j.1468-0068.2010.00800.x.

King, D. 2014. *The Commissar Vanishes: The Falsification of Photographs and Art in Stalin's Russia New Edition*. London: Tate.

Koopman, C. 2019. *How We Became Our Data: A Genealogy of the Informational Person*. Chicago: University of Chicago Press.

Krücken, G., and G. S. Drori, eds. (2010). *World Society: The Writings of John W. Meyer,* 1st ed. Oxford: Oxford University Press.

Lackey, J. 2010. *Learning from Words: Testimony as a Source of Knowledge*. Oxford; New York: Oxford University Press.

Langton, R. 1993. "Speech Acts and Unspeakable Acts." *Philosophy & Public Affairs* 22 (4): 293–330. https://www.jstor.org/stable/2265469.

Langton, R., and J. Hornsby. 1998. "Free Speech and Illocution." *Legal Theory* 4 (1): 21–37. https://doi.org/10.1017/S1352325200000902.

Lash, S. M. 2002. *Critique of Information*. London: SAGE Publications Ltd.

Loiperdinger, M., and B. Elzer. 2004. "Lumiere's Arrival of the Train: Cinema's Founding Myth." *The Moving Image* 4 (1): 89–118. http://doi.org/10.1353/mov.2004.0014.

MacKinnon, C. A. 1987. *Feminism Unmodified*. Cambridge: Harvard University Press.

Mills, C. W. 2017. *Black Rights/White Wrongs: The Critique of Racial Liberalism*. New York, NY: Oxford University Press.

Mills, C. W. 2007. "White Ignorance." In *Race and Epistemology of Ignorance*, edited by S. Sullivan, and N. Tuana, 13–38. Albany, N.Y: State University of New York Press.
Origgi, G. 2017. *Reputation: What it Is and Why it Matters,* S. Holmes, and N. Arikha (translators). Princeton: Princeton University Press.
Orwell, G. 1961. *1984*. New York City: Signet Classic.
Peters, J. D. 1988. "Information: Notes toward a Critical History." *Journal of Communication Inquiry* 12 (2): 9–23. https://doi.org/10.1177/019685998801200202.
Rawls, J. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.
Rawls, J. 2001. *Justice as Fairness: A Restatement,* edited by E. Kelly, 2nd ed. Cambridge, Mass: Belknap Press.
Rini, R. 2020. "Deepfakes and the Epistemic Backstop." *Philosophers Imprint* 20 (24): 1–16. https://quod.lib.umich.edu/p/phimp/3521354.0020.024/1.
Risse, M. 2020. *On Justice: Philosophy, History, Foundations*. New York: Cambridge University Press.
Rosenberg, D. 2013. "Data before the Fact." In *"Raw Data" Is an Oxymoron*, edited by L. Gitelman, 15–40. Cambridge, Massachusetts; London, U.K.: The MIT Press.
Sanders, B., and I. Illich. 1989. *ABC: Alphabetization of the Popular Mind*. New York: Vintage.
Savedoff, B. E. 2000. *Transforming Images: How Photography Complicates the Picture*. Ithaca, NY: NCROL.
Smith, D. W. 2013. *Husserl*. London; New York: Routledge.
Snyder, T. 2019. *The Road to Unfreedom: Russia, Europe, America*. New York: Tim Duggan Books.
Susskind, J. 2018. *Future Politics: Living Together in a World Transformed by Tech*. Oxford; New York: Oxford University Press.
Tegmark, M. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Knopf.
Walton, K. 1984. "Transparent Pictures: On the Nature of Photographic Realism." *Nous* 18 (1): 67–72. https://doi.org/10.2307/2215023.
Watkin, C. 2018. *Michel Foucault*. Phillipsburg, New Jersey: P & R Publishing.
Watson, L. 2018. "Systematic Epistemic Rights Violations in the Media: A Brexit Case Study." *Social Epistemology* 82 (2): 88–102. https://doi.org/10.1080/02691728.2018.1440022.
Wells, H. G. 2016. *World Brain*. Redditch, Worcestershire: Read Books Ltd.
Wenar, L. 2003. "Epistemic Rights and Legal Rights." *Analysis* 63 (2): 142–6. https://doi.org/10.1111/1467-8284.00024.
Wenar, L. 2015. "Rights." In *Stanford Encyclopedia of Philosophy*. Also available at https://plato.stanford.edu/archives/fall2015/entries/rights/ (accessed October 9, 20).
Wrone, D. R. 2003. *The Zapruder Film: Reframing JFK's Assassination*. Lawrence, Kan: University Press of Kansas.
Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.