

# Deepfakes and the Epistemic Backstop

Regina Rini (York, Toronto) · rarini@yorku.ca

**Abstract:** Deepfake technology uses machine learning to fabricate video and audio recordings that represent people doing and saying things they've never done. In coming years, malicious actors will likely use this technology in attempts to manipulate public discourse. This paper prepares for that danger by explicating the unappreciated way in which recordings have so far provided an epistemic backstop to our testimonial practices. Our reasonable trust in the testimony of others depends, to a surprising extent, on the regulative effects of the ever-present possibility of recordings of the events they testify about. As deepfakes erode the epistemic value of recordings, we may then face an even more consequential challenge to the reliability of our testimonial practices themselves.

This is a working draft as of June 2019. It is currently under journal review. You are welcome to circulate this manuscript. If citing, please indicate the June 2019 manuscript draft, available on PhilArchive.

Deepfakes are fabricated video or audio recordings created through machine learning technology. A computer program uses a large data set of real recordings to build a model of the facial/vocal characteristics of a person, then superimposes this onto recordings of another person. The effect is an *apparent* recording of a well-known person doing or saying something they never did. If you haven't seen a deepfake, you should probably stop and watch one before continuing with this paper; it's a phenomenon that's hard to grasp without ostensive demonstration.<sup>1</sup>

The current output isn't entirely convincing. But that may change, and soon. Which means we should start asking: what could happen to our collective, democratic processes of information-sharing and debate when our leaders can be deepfaked into doing or saying whatever a malicious agent desires?

---

<sup>1</sup> For an especially entertaining example, you can watch 'Barack Obama' insult Donald Trump, in a demonstration from BuzzFeed and Jordan Peele. See: Mack, D. (2018, April 17). This PSA About Fake News From Barack Obama Is Not What It Appears. *Buzzfeed*. Retrieved from <https://www.buzzfeednews.com/article/davidmack/obama-fake-news-jordan-peeel-psa-video-buzzfeed#.noWQRo6Em>

Hints of this worry already appear in journalistic discussions, but there's been very little published by academics.<sup>2</sup> To my knowledge, there is nothing yet from philosophers on the subject. So consider this essay an exercise in prophylactic political epistemology; I aim to raise worries about deepfakes and the erosion of knowledge in democratic societies *before* it begins to happen.

### **Testimony, Recordings, and the Epistemic Backstop**

I'll come back to deepfakes shortly. First, we need to get clear on the epistemic environment in which they operate. I'm going to argue that audio and video recordings currently serve a distinctive role in public discourse. Specifically, they regulate our testimonial practices, providing what I'll call an *epistemic backstop*.

Our collective epistemic practices are highly reliant on testimony, the transmission of knowledge via say-so. When a credible person tells me that she has seen or heard something, I am typically justified in believing the contents of her claim simply on the basis of her testifying to it. Of course, this justification can be defeated by evidence against the proposition, or undercut by evidence that my interlocutor is unreliable. But absent those factors, I seem to have a default justification to accept testimony as evidence, and in many cases to adopt a belief simply on testimonial recommendation. Since any individual knower only has so much time and cognition available, our knowledge of the world would be severely impoverished if we could not rely upon testimony.<sup>3</sup>

---

<sup>2</sup> See Maher, S. (2018, December 28). Fake Video is a big problem. And it's only going to get worse. *MacLean's*. Retrieved from <https://www.macleans.ca/opinion/fake-video-is-a-big-problem-in-2019-it-gets-worse/>. For a legal academic perspective, see Chesney, Robert and Danielle Keats Citron Forthcoming 2019,

<sup>3</sup> The philosophical literature on testimony has grown significantly in the last three decades. For important examples, see Coady (1992), Lackey (2008), and Goldberg (2010).

Of course, some philosophers – being philosophers, after all - have skeptical worries about our reliance on testimony. But most theorists accept some form of justification via testimony, and all socially functional humans rely on it in practice, whatever their theoretical qualms. A major part of the rationality of this practice comes from widespread awareness of enforceable *testimonial norms*. When a person attempts to provide testimony, she is taken to be implying that she is both *sincere* and *competent* about the matter on which she testifies. Put simply: she really means it, and she knows what she’s talking about.<sup>4</sup> A person who regularly transmits misinformation violates one or both of these norms and will likely acquire a reputation as a liar or an idiot. Fear of those reputational consequences gives everyone – even those lacking in a good will - a reason to cooperate with the norms. This excerpt from Michael Cohen’s February 2019 testimony before the US House Oversight Committee provides a remarkably succinct illustration of *both* the ways that bad testimonial practice can blemish one’s reputation:

Rep. Comer: You called Mr. Trump a cheat. What would you call yourself?

Michael Cohen: A fool.<sup>5</sup>

So far these are just familiar philosophical theses about the nature of testimony. But now I want to draw attention to the role of *recordings* in regulating our testimonial practices, something rarely appreciated by analytic epistemologists. Neglect of this role has perhaps been reasonable until now, but deepfakes will soon make it dangerous.

Here is the main idea: video and audio recordings function as an *epistemic backstop*. They regulate our testimonial practices, sometimes bluntly but more often subtly. The availability of

---

<sup>4</sup> Here I am synthesizing (and simplifying) convergent points from two different philosophical literatures: the knowledge norm of assertion (Williamson 1996, Lackey 2007) and the nature of interpersonal trust (Baier 1986; Jones 1996).

<sup>5</sup> See Prasad, R. & Matza, M. (2019, February 27). Reaction to Cohen testimony. BBC. Retrieved from <https://www.bbc.com/news/live/world-us-canada-47390920/page/4>

recordings undergirds the norms of testimonial practice, increasing the incentive of testifiers to speak with sincerity and competence. Our awareness of the possibility of being recorded provides a quasi-independent check on reckless testifying, thereby strengthening the reasonability of relying upon the words of others. Recordings do this in two distinctive ways: actively correcting errors in past testimony and passively regulating ongoing testimonial practices. I will take these in turn.

*Acute correction* is the active role recordings play in regulating testimony. Think of familiar ‘let’s check the tape’ scenarios: someone has testified that an event transpired in such-and-such, but then a surreptitious recording reveals otherwise. Perhaps the most famous example is Richard Nixon’s ‘smoking gun’ tape. In 1974, Nixon denied for months that he’d had any foreknowledge of the Watergate coverup. His aides were dispatched to testify to the public that the president was innocent; they became the outward links in a testimonial chain running back to Nixon himself. Unfortunately for the president, his historical grandiosity had led him to install a secret audio recording system in the Oval Office. After much dissembling and legal evasion, Nixon turned over a June 1972 recording on which he could be clearly heard ordering the CIA to interfere in the FBI’s Watergate inquiry. The effect was devastating; Nixon was revealed to have unambiguously flouted the sincerity norm of testimony, and his aides further down the testimonial chain were left out in the cold. This last fact seems to have been especially important, as many of the Nixon true-believers now felt personally betrayed by a president exploiting their testimonial assistance. Even his innermost defenders turned on him. Within weeks, the president resigned.<sup>6</sup>

You might think modern politicians have learned from examples like this. But recordings still provide acute correction in politics. In spring 2018, Junichi Fukuda, a senior official in the Japanese Finance Ministry, was accused of sexually harassing a reporter during an interview. Fukuda denied the

---

<sup>6</sup> Woodward and Bernstein (1976).

allegations - but it then emerged that the reporter had pressed the audio record button on her mobile phone. When she publicly released this recording, Fukuda continued to deny the allegations. He admitted the voice in the recording perhaps *sounded* like him, but claimed “I cannot tell since I can only hear my voice through my own body.” The public was not convinced. Fukuda was quickly forced to resign.<sup>7</sup>

Recording can also provide acute correction in cases of conflicting or confused testimony. An extremely famous example is the Zapruder film of the 1963 assassination of John Kennedy. In this case, there were thousands of independent eyewitnesses, an entire crowd in a large public location. But people were sure they witnessed different things – a gunshot echoing from here, another shot from there. Complexities of speed, emotion, distance, and memory made it hard to know which eyewitness testimony to trust. The Zapruder film provided an independent check on all the testimonial noise, allowing investigators to calibrate a single narrative (though, of course, not everyone was convinced).<sup>8</sup>

As important as these acute corrections are, I suggest that the passive regulation role of recordings is still more important, if less fully appreciated. This is the idea: part of the reason our ordinary testimonial practice allows us to trust one another to be sincere and competent is that we all know that, at any time, we *might* be within the range of an audio or video recorder. This is especially true for public figures, who can expect to acquire the interest of smartphone equipped observers in every room they enter. Your background awareness that the event you are testifying about *might* have

---

<sup>7</sup> Yamaguchi, M. (2018, April 18). Japan finance official denies sexual misconduct allegation. *The Seattle Times*. Retrieved from <https://www.seattletimes.com/nation-world/japan-finance-official-denies-sexual-misconduct-allegation/>

<sup>8</sup> The Zapruder film also provides instruction on the risks of *over*-reliance on recordings. According to a modern theory by journalist Max Holland, the Warren Commission was misled by assuming that the Zapruder tape covered the entire assassination, when in fact the first gunshot (which missed) was fired before Zapruder turned on his camera. Trying to interpret all three shots within the film’s short timeframe led to factual inconsistencies that were quickly seized upon by conspiracy theorists. See Max Holland, ‘The Truth Behind JFK’s Assassination’ (2014) *Newsweek* November 20, 2014. <https://www.newsweek.com/2014/11/28/truth-behind-jfks-assassination-285653.html> [accessed Feb 27 2019]

been recorded is a good reason to be as sincere and competent as possible, even if you lack characterological honesty or aptitude.

Of course, that's not always enough to keep people from lying or overstepping their competence. People still do it, as the Fukuda case shows. But I suspect that our tendency to mis-testify has declined significantly since the arrival of widespread recording. We can see this passive regulatory effect operating in some complex examples played out in national political life. In May 2017, US President Donald Trump fired FBI Director James Comey, who alleged that this was a result of his refusal to help the president suppress an investigation of Russian interference in the 2016 presidential election. When this made headlines, Trump tweeted: "James Comey better hope that there are no "tapes" of our conversations before he starts leaking to the press!"<sup>9</sup>

This was a rather insidious exploitation of the testimony-regulating function of recordings. Trump was implicating that, like Nixon, he keeps a secret recording device in the Oval Office, and that he had tapes showing Comey's testimony to be false. But no such tapes ever emerged. In fact, this instance shows how the regulatory function can cut both ways; the press immediately demanded that Trump release the hypothetical recordings. Comey asked him to do so as well. A month later, the twittering president followed up: "With all the recently reported electronic surveillance, intercepts, unmasking and illegal leaking of information, I have no idea ... [second tweet] whether there are "tapes" or recordings of my conversations with James Comey, but I did not make, and do not have, any such recordings."<sup>10</sup>

---

<sup>9</sup> @realDonaldTrump. (2017, May 12). Retrieved from <https://twitter.com/realdonaldtrump/status/863007411132649473>

<sup>10</sup> @realDonaldTrump. (2017, June 22). Retrieved from <https://twitter.com/realdonaldtrump/status/877932907137966080> and @realDonaldTrump. (2017, June 22). Retrieved from <https://twitter.com/realdonaldtrump/status/877932956458795008>

This is a remarkable case for seeing the multiple roles of recordings in testimonial practice.

When Trump was unable to produce the “tapes”, many observers took this as a charge against his own credibility, and further support for Comey’s testimony. After all, the Director of the FBI was surely aware of the Nixon precedent for a self-bugged Oval Office, and presumably experienced passive regulation of his testimony by the thought that tapes later might emerge. Trump’s incompetent insinuation of falsifying tapes instead reminded the public that Comey had a strong reason to be telling the truth.

As these examples show, recordings (or the ongoing possibility thereof) play several important roles in supporting the rationality of our testimonial practices. It is surprising, then, that epistemologists have not much attended to these roles. I suspect this is because analytic philosophers prefer to focus on necessary or conceptual conditions for knowledge transmission, the sort of factors you could find at work in the epistemic practice of any era or any society, unlike the historically contingent effects of recordings. The first known audio recording was made in France in 1857.<sup>11</sup> The oldest extant visual recording, *Roundhay Garden Scene*, dates to 1888.<sup>12</sup> A phenomenon that began to exist only 150 years ago is unlikely to attract the attention of a perspective attuned to universal logical constants.

But this is a misleading way of thinking. Though recordings are historically contingent, their role in contemporary social epistemic practice is pervasive. Recording technology is older than any currently existing person. Everyone living today (at least in wealthy countries) has *always* lived with the implicit understanding that recordings provide a check on disputed testimony. Our social epistemic practices

---

<sup>11</sup> Flatow, I. (2008, April 4). 1860 'Phonautograph' Is Earliest Known Recording. *National Public Radio*. Retrieved from <https://www.npr.org/templates/story/story.php?storyId=89380697>

<sup>12</sup> Smith, I. (2016, January 10). “Roundhay Garden Scene” recorded in 1888, is believed to be the oldest surviving film in existence. *The Vintage News*. Retrieved from <https://www.thevintagenews.com/2016/01/10/roundhay-garden-scene-is-believed-to-be-the-oldest-known-video-footage/>

have adapted to five generations' awareness of this possibility. As a result, the epistemic role of recording can be diaphanous; it is hard to see a lens that was put in place before you were born.

I suggest, then, that the epistemic role of recordings has fallen into a characteristic gap in analytic philosophers' attention. Somewhere between the logically universal and the suddenly emergent lies the realm of historically entrenched social practice. As a discipline, analytic philosophy has tended to ignore that realm.<sup>13</sup> But this becomes dangerous when social conditions suddenly change. What was diaphanous while intact may become disorientingly opaque once torn. That is what I fear may soon happen to the epistemic role of recordings. To see this danger, we need to understand deepfakes.

### **Deepfakes and public deception**

Deepfakes came to public attention in December 2017, following an article in *Motherboard* by tech writer Samantha Cole.<sup>14</sup> As with so much on the internet, pornography was the main vector. The basic idea is this: if you feed hundreds of hours of a celebrity's video appearances into a machine learning algorithm, the algorithm will build an adaptable model of their face, which can then be digitally inserted over the face of a different person in a different recording. So, if you have a desire to project your favorite actress into your favorite pornographic scene, the technology can make your (invasive, creepy) dream come true. Connoisseurs of the form gathered on the internet discussion forum Reddit,

---

<sup>13</sup> Continental philosophers, of course, are much more comfortable in the land of historically entrenched social practice. A huge literature, of obvious relevance to the topic of this paper, opens out from Walter Benjamin's *The Work of Art in the Age of Mechanical Reproduction* (1935). I am unaware of anything in continental work touching *directly* on my argumentative claims, though it seems best to confess that my own training is mostly analytic. In genuine social-epistemic fashion, I would be pleased if a reader with better knowledge of other traditions pointed me to a close parallel.

<sup>14</sup>

Cole, S. (2017, December 11). AI-Assisted Fake Porn is Here and We're All Fucked. *Motherboard*. Retrieved from [https://motherboard.vice.com/en\\_us/article/gydydm/gal-gadot-fake-ai-porn](https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn)

where they swapped source images and fabricated outputs. One user - whose Reddit handle 'deepfakes' came to stand for the phenomenon itself - provided a free software tool ('FakeApp') allowing anyone with a decent home computer to make their own videos. Following media attention, Reddit banned the deepfake community for violating its 'involuntary pornography' rules.<sup>15</sup> Of course, the internet is a wide and wild place, and deepfake pornography is far from eradicated.

Obviously, the use of deepfakes to present people doing and saying things they never did - especially in pornography - leads to serious ethical worries.<sup>16</sup> But in this paper I will focus on the risks of epistemic mischief. The most obviously dangerous applications are in politics. Several journalistic entities have already published political deepfakes, featuring e.g. Donald Trump's face superimposed on Angela Merkel's body or Barack Obama appearing to call Trump a "total and complete dipshit". In May 2018 the Flemish Socialist Party posted a deepfake video appearing to show Trump urging Belgium to withdraw from environmental treaties. The Party later claimed that it meant the video to be only a provocation to conversation, not to fool anyone into believing its content. Yet some internet commenters appear to have taken the video as real.<sup>17</sup>

In January 2018, technologist John Wiseman documented what may be the first attempted deepfake with a political motive. The request was posted to the Reddit deepfakes forum (before the

---

<sup>15</sup> Robertson, A. (2018, February 7). Reddit bans 'deepfakes' AI porn communities. *Verge*. Retrieved from <https://www.theverge.com/2018/2/7/16982046/reddit-deepfakes-ai-celebrity-face-swap-porn-community-ban>

<sup>16</sup> I discuss the ethics of deepfakes elsewhere. [unpublished paper with Leah Cohen]

<sup>17</sup> See Der Burchard, H. V. (2018, May 21). Belgian socialist party circulates 'deep fake' Donald Trump video. *Politico*. <https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video/>; Silverman, C. (2017, April 17). How To Spot A Deepfake Like The Barack Obama–Jordan Peele Video. *Buzzfeed*. Retrieved from <https://www.buzzfeed.com/craigsilverman/obama-jordan-peepe-deepfake-video-debunk-buzzfeed>

forum was banned) and does not appear to have been fulfilled. This is the text as presented in a screencap Wiseman posted to Twitter:

In Russia we have a big problem with gay activist rights. The governor of Chechnya (region in Russia) Ramzan Kadyrov supposedly ordered the execution of 200+ gay people in his region. Vladimir Putin (our president) is a big pussy and doesn't want to do anything with him. Can someone please make a gay video with Ramzan and some other random guy? We can get it viral within his region. Whatsapp channels with 100K+ subscribers will go crazy over it. Everyone will know it's fake, but his reputation will be done. (but not Ramzan with Putin together, because it would just be considered stupid)<sup>18</sup>

This was a request to utilize the technology for a political goal, to try to undermine a politician. (In this particular case the politician is rather loathsome, but of course there is nothing confining the use of deepfakes to laudable political objectives.) And the example points to something very important. Notice that the Reddit user says, "everyone will know it's fake, but [Kadyrov's] reputation will be done." The thought is that a deepfake video needn't be positively *believed* by all viewers in order to be effective. Merely getting a suggestive video into public distribution may be enough, even if many viewers realize it is fake.<sup>19</sup>

Still, we might ask: just how convincing *is* this technology? If you look at current examples on the web, especially those by Reddit amateurs, you might be dubious. Many deepfake videos have giveaway digital artifacts. The edges of the jaw are blurred unnaturally. The eyes don't blink quite as you'd expect. Though you might not notice these things while watching quickly on a low-resolution platform, you can see them if you pay attention. A critical viewer, with access to a large digital platform, might

---

<sup>18</sup> /r/deepfakes post by Reddit user 'ethan1el', as embedded in January 25 2018 tweet by John Wiseman (@lemonodor). <https://twitter.com/lemonodor/status/956652112678551552> [[accessed March 4 2019]]

<sup>19</sup> This point is also shown by what *might* be the first case of an actual political deepfake. In late 2018, Gabonese president Ali Bongo appeared in a video that his opponents claim is machine-generated. Bongo suffered a stroke months earlier and had not been seen in public since. It is still unclear whether the video was a deepfake, but the Gabonese political scene was rattled either way. See Ali Breland, 'The Bizarre and Terrifying Case of the "Deepfake" Video that Helped Bring an African Nation to the Brink', *Mother Jones* March 15, 2019; <https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/> .

spread a debunking nearly as quickly as the deepfake itself.<sup>20</sup> So perhaps deepfakes won't be very successful political tools.

The problem with this hope is that, of course, new technology eventually improves. To understand what future deepfakes will be like, we shouldn't look at Reddit amateurs. We should look at professional computer scientists. In 2016, a group of researchers in Germany and California presented a technique called 'Face2Face', which uses an algorithmic process to impose a famous person's visage over an actor's *in real time*. You can watch a demonstration online: an actor sits in front of a camera, moving his face through various contortions. At the same time, the computer screen displays the face of George W. Bush doing the same things.<sup>21</sup> The technology isn't perfect (Bush's face does look a bit odd), but again this is still early days. More recently, some of the same researchers proposed a 3-D modeling technique doing the same thing with fabricated videos of a person's entire body - similar to Hollywood motion capture technology.<sup>22</sup>

You might be tempted to think that fabricated video is of little importance without corresponding audio. Even if a malefactor can make Barack Obama's face do whatever they want, we'll be able to tell that it's not speaking with Obama's voice. (The Obama deepfake mentioned earlier employs noted mimic Jordan Peale to provide the voice. Not everyone will have access to such effective vocal talent!) Unfortunately, there are deepfake technologies for voice as well. Researchers at Princeton and Adobe (the makers of Photoshop) debuted a technique called 'VoCo' in 2017. It allows the user to alter the content of a spoken audio recording simply by typing new words. The algorithm synthesizes

---

<sup>20</sup> Though the short history of fake news suggests otherwise. Fake stories spread faster and farther than true ones, presumably including their own debunkings (Vosoughi, Roy and Aral 2018).

<sup>21</sup> 'Face2Face: Real-time Face Capture of Reenactment of RGB Videos (CVPR 2016 Oral)'. <https://www.youtube.com/watch?v=ohmajJTcpNk> . See also the project page at [https://web.stanford.edu/~zollhofer/papers/CVPR2016\\_Face2Face/page.html](https://web.stanford.edu/~zollhofer/papers/CVPR2016_Face2Face/page.html)

<sup>22</sup> Liu et al. (2018).

what the speaker's voice would have sounded like with the altered phonemes. It's as if you can make any voice read from any script you'd like.<sup>23</sup>

You might now say: okay, this technology is scary, but won't it won't be cancelled out by equally powerful detection technology? Perhaps we can turn machine learning loose on the internet and have it diagnose characteristic patterns in manipulated recordings.<sup>24</sup> There are two problems with this thought. The first is technical. Even if counter-deepfake technology temporarily overcomes fakery, this may just be the first move in a machine learning arms race, where fakers continually change strategies to stay a bit ahead of detection. Second, and more fundamentally, there is a social epistemic problem. Even if we had reliable deepfake detection technology, past experience with fake news suggests that corrections rarely travel as far as initial fakes and are often not as readily believed. In any case, this offers at best a kind of epistemic chaos, with corrections chasing vivid fakes around the internet, being variously championed by partisan communities.

Another technical proposal is called watermarking. This involves inserting a layer of invisible-to-the-naked-eye data into recordings, which becomes computer-detectably distorted even when the visible data is expertly faked. The problem with this proposal is that it requires access to the original recording device's intact watermark. This is strong protection for specialized recording devices (e.g. the CCTV cameras at Fort Knox), but won't work for random smartphone footage circulating on the internet. This could be a general solution only with the establishment of an international database of registered device watermarks, which seems impractical to say the least.<sup>25</sup>

---

<sup>23</sup> Jin et al (2017).. You can see a demonstration at: <https://www.youtube.com/watch?v=RB7upq8nzIU>

<sup>24</sup> The computer scientist Hany Farid is a leader on this research. See Schwartz, O. (2018, November 12). You thought fake news was bad? Deep fakes are where truth goes to die. *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth>

<sup>25</sup> For research on watermarks, see Quring, Arp and Rieck (2017).

Perhaps most worrisomely, we can only see the publicly available deepfake research done by universities and corporations. One must assume that several extremely well-funded national intelligence services are also hard at work on their own in-house versions of deepfakery. For all we know, their technology is already substantially improved, and they are simply waiting for the right moment to deploy it. (Would we know it if they already *had*?) In any case, the point of this paper is to think ahead, perhaps five years or ten, to the effects of an improved version of the technology – to allow us to begin preparing for the epistemic repercussions before they become dangerously real.

### **Backstop crises**

Let's return to the concept of the epistemic backstop to see, now, why deepfakes are so dangerous. I said earlier that recordings perform acutely corrective and passively regulative roles in our testimonial practices. I'll now argue that the increasing availability of deepfake technology – and especially growing public *awareness* of its existence – will severely disrupt the functioning of the epistemic backstop.

Think ahead to a day when deepfake technology is convincing and widely available. The problems, I suggest, will start with events I call *backstop crises* – moments when the corrective and regulative functions of recordings are made salient, but then quickly undercut by the spectre of deepfakery. Imagine Richard Nixon had said: “look, that wasn't me on the smoking gun tape. They used that VoCo technology to make it sound like me ordering CIA interference. But it wasn't!” Of course, that would not have been a very plausible claim in 1974. But now imagine that at the start of the 2020 US presidential campaign, an audio recording emerges which certainly sounds like Donald Trump colluding with Russian intelligence operatives. Suppose Trump insists that it wasn't him, that he has been deepfaked into an entirely falsified conversation.

I don't know how many people would believe Trump, or how many would believe the recording. I suspect views would break along predictably partisan lines. But I can say this: though not myself a fan of Donald Trump, I would have serious doubts about the veracity of this tape. Even if you have strong prior reasons to suspect that Trump *did* collude with Russian intelligence, awareness of deepfake technology furnishes reason to doubt this supposed piece of evidence. After all, many political actors, domestic and foreign, would have strong motives to create such a tape, regardless of its underlying truth.

And the point of this example doesn't require that everyone *believe* a deepfaked tape. The point is the public controversy itself. What makes a backstop *crisis* is the sudden public realization that there is no longer any such thing as a "smoking gun" tape. Regardless of how we resolve our first majorly public encounter with deepfakes, we will all come away with a growing sense of displaced epistemic reality. Very quickly, video and audio recordings may lose their status as acute correctors of the testimonial record. Once everyone has seen that a supposedly authoritative recording can be reasonably challenged, we will all start to wonder about the next recording, and the next, and the next...

Here is another scenario to make the idea more vivid. Imagine a politician skyping in to a tv news show and, under pressure from interviewers, saying something foolish or offensive. Instantly, the pundits decree this a deadly gaffe, the beginning of the end of the politician's career. But then the politician's staffers claim that this wasn't the politician at all. They assert that their politician was nowhere near a camera at the time. Rather, it must have been an actor skyping in, using real-time deepfake technology like Face2Face to steal the image and voice of the poor misrepresented politician!

Imagine the fallout from such a spectacle. Hours and hours of cable news fulmination over the plausibility of a guest's appearance being 'hacked'. Linguists and computer scientists summoned to provide rival millisecond breakdowns of the recording. 'See the lip move like that, there? That can't be

real!’ Or: ‘Look, that partial blink is very hard to fake with an algorithm. This is the real thing.’ Imagine this going for days, or weeks, with no clear resolution. In the end, presumably, public opinion will break along partisan lines. If you like the politician, you’ll believe the denial. If you don’t, you’ll allege that they are crying deepfake wolf, and hate them all the more for it.

Regardless of the outcome – whether we believe the politician or the video - the process *itself* is a backstop crisis. We will all confront a suddenly plausible skepticism about the knowledge-bearing potential of video and audio, after lifetimes relying on them as solid testimonial anchors. This has the makings of an epistemic crisis on the order of beginning to suspect that a hallucinogen has been pumped into the city’s water supply.

A similar sort of crisis may come from conflicting evidence. When many witnesses attest to differing versions of the same event, we often turn to smartphone recordings to decide what really happened. In January 2019 – when America media consumers were briefly obsessed by the details of an unhappy encounter between a Native elder and a teenage Trump fan on the steps of the Lincoln Memorial – we knew already to search the internet for multiple videos, shot from different angles at different times. Yet even a wealth of documentary evidence was insufficient to settle our collective Rashomon interpretation of the smirk-or-not on a young man’s face.<sup>26</sup> Imagine that one or two videos appeared to show an otherwise undocumented moment of blatant hatred – a slur escaping the lips, for instance. Then imagine the accusation that these supposedly decisive recordings are deepfakes. Or, instead, it was the others, the ones that seemed to show the opposite! The viewer may select their own targets of skepticism, as the partisan spirit desires.

---

<sup>26</sup> Christine Emba (2019). ‘What a dead samurai can teach us about the Covington controversy’. *Washington Post* January 24, 2019. [https://www.washingtonpost.com/opinions/what-a-dead-samurai-can-teach-us-about-the-covington-controversy/2019/01/24/eeddee12-201d-11e9-8e21-59a09ff1e2a1\\_story.html](https://www.washingtonpost.com/opinions/what-a-dead-samurai-can-teach-us-about-the-covington-controversy/2019/01/24/eeddee12-201d-11e9-8e21-59a09ff1e2a1_story.html)

Earlier I stressed that the acute corrective role of recordings is less important than their passive regulatory role. We can see that most clearly now, as we try to imagine the long-term effects of highly public backstop crises. Once we know that recordings can be deepfaked – and that *veridical* recordings can be convincingly dismissed – our motivation to be responsible testifiers may slowly erode. We would no longer worry that someone might have recorded the events about which we are testifying. If the recording goes against us, we can always cry ‘deepfake!’ And we might even be right.

This is the real danger of deepfakes: not that they will trick us into believing false content, but that they will gradually eliminate the epistemic credentials of *all* recordings, to an extent that video and audio no longer serve their passive regulative function in testimonial practice. As that happens, the reasonableness of expecting testifiers to be (usually) sincere and competent will begin to diminish. Within a few years, we may have little reason to trust the testimony of strangers, as the norms securing their anticipated cooperation come gradually undone. Backstop crises triggered by deepfakes may be only the harbingers of a slow-boiling but deeply consequential epistemic maelstrom.

### **More about the epistemology of recordings**

Maybe you aren’t so worried. Maybe you think: look, we’ve been through this before. We used to be able to trust photographs. Then along came clever editing, and skilled airbrush artists, and finally Photoshop. Now everyone knows that everyone can manipulate photographs, but there’s been no epistemic catastrophe. So we’ll adjust to deepfakes too.

This is a reasonable point, which I’ll address directly in the next section. To assess the objection, we need to say a bit more about the epistemic underpinnings of recordings themselves. When a recording *does* provide epistemic access to its subject, what sort of epistemic role is it playing? As I noted earlier, analytic philosophers have had surprisingly little to say about video and audio recordings.

But there is a significant literature on the epistemology of still photographs. We might use this to build up an account of recordings.

Much of this literature is framed in response to Kendall Walton's famous *transparency* thesis, which holds that photographs enable direct, literal perception of their objects.<sup>27</sup> Suppose I want to learn about Queen Victoria's dog Looty, the supposed ancestor of all Pekingese in Britain and North America. (Looty was so named because she had been stolen – along with many equally priceless objects - from the Chinese Imperial Summer Palace by rampaging British troops during the Second Opium War.) First I look at an oil painting of Looty, executed by Friedrich Wilhelm Keyl in 1861. Then I look at a photograph of Looty, grumpy on an ornate stiff-backed chair, taken by William Bambridge in 1865.<sup>28</sup> According to Walton, in the latter case I am literally *seeing* Looty, across time via this photograph, just as surely as if I were seeing her across space via a telescope. By contrast, when I look at the painting, I am not literally seeing Looty, only a pictorial representation of her, mediated through Keyl's perceptions and artistic expression.

Walton's proposal has been extremely controversial, chiefly due to the oddity of implying that right now I am literally seeing a dog who died 140 years ago.<sup>29</sup> But the same epistemic point comes through in other ways. Robert Hopkins shifts from the idea that I literally see Looty to the slightly more arcane suggestion that a photograph can place me in a necessarily veridical *seeing-in* relation to Looty. Assuming the photograph was made and reproduced according to standard photographic norms, then

---

<sup>27</sup> Walton 1984

<sup>28</sup> Both the painting and photograph are held by the Royal Collection. You can see them online at, respectively, <https://www.rct.uk/collection/406974/looty> and <https://www.rct.uk/collection/2105644/looty-the-pekingese> . To learn more about Looty herself, see Cynthia Haven (2010), 'Stanford historian tells why the West rules – for now'. *Stanford News* September 14, 2010. <https://news.stanford.edu/news/2010/september/morris-west-rules-091410.html>

<sup>29</sup> For discussion, see Cohen and Meskin 2004.

my experience of seeing Looty in the picture is roughly as reliable as seeing Looty with my own eyes.<sup>30</sup>

I'm not literally seeing Looty, but (assuming there were no shenanigans in the causal pathway from Bambridge's shutter to my screen) I gain the same epistemic access to Looty as if I were.

Dan Cavedon-Taylor develops this idea further, claiming the distinct advantage of photographs over hand-made pictures is that the former generate *perceptual knowledge*, while the latter can only support *testimonial knowledge*. When I look at Keyl's painting of Looty, I am relying upon the painter's skill and honesty in pictorial representation just as much as I would rely upon his competence and sincerity in a verbal testimonial report. By contrast, Cavedon-Taylor argues, Bambridge's photo is not mere testimonial evidence; it belongs in the same category as ordinary perceptual experience, with the same epistemic immediacy. He says: "when we see a photograph that depicts *x* as *F*, say, our default doxastic response is to believe that *x* is *F* – and to only withhold assent if we possess reasons *against* thinking the photograph creditworthy."<sup>31</sup>

One more example will illustrate this contrast. Leo von Klenze's 1857 painting *The Temple of Concordia at Agrigento* depicts the eponymous Sicilian ruin. When I saw this painting hanging in the Alte Nationalgalerie in Berlin, I was confused. I've been to Agrigento, and something about the painting seemed wrong to me, though I couldn't articulate just what. Later I looked up a photo of the temple and understood: the viewing angle Klenze depicts is not actually possible; the temple and the city simply are

---

<sup>30</sup> Hopkins 2012. This is a slightly compressed version of a nuanced view. One worry about Hopkins' view is that he appears to concede that photographs are only reliable when viewers see in the photo "only facts consistent with *perfect* general knowledge of how things are" (719). Given the difficulty of attaining second-order knowledge that one's judgements cohere with perfect general knowledge, I worry that Hopkins' proposal encourages practical skepticism regarding *all* real world photographic experiences. But I'll leave that thought to the side here.

<sup>31</sup> Cavedon-Taylor 2013, 294.

not arranged in that way. Klenze seems to have taken artistic license with perspective to provide more pleasing composition.<sup>32</sup>

This example shows the different epistemic powers of paintings and photographs. My pictorial experience of the temple as in Klenze's painting was at odds (inchoately) with my perceptual memory. I was unable to resolve this tension until I examined a photo of the same ruins, which quickly decided me in favor of my memory and against the painting. This is because, just as Cavedon-Taylor suggests, paintings provide only testimonial evidence, which is typically trumped by perceptual evidence of the sort photographs provide. Because I know that Klenze's aesthetic sensibilities play a crucial role in his depiction of the temple – and this is not typically true of photographs – I treat the painting with the same merely provisional authority as testimony. As Cavedon-Taylor puts 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” (288-9).

Plausibly, all of the above is just as true for audio and video recordings as it is for still photographs. Recordings provide us with a form of perceptual evidence, which enjoys a stronger presumptive authority than testimonial evidence. And this is exactly why recordings are so well suited to provide the epistemic backstop. Their stronger evidential weight allows them to provide acute correction of deviant testimonial practices and passive regulation of trustworthy testimonial norms.

We can now recast the points of the preceding sections more precisely. The danger posed by deepfakes is that successive backstop crises will gradually transform our attitudes toward the epistemic status of recordings. In effect, recordings will be demoted from sources of perceptual evidence to sources of mere testimonial evidence. And if they are simply just another source of testimony, then they

---

<sup>32</sup> You can see the painting here: <https://www.smb.museum/en/exhibitions/detail/concordia-kunst-und-wissenschaft-in-eintracht.html> . Gallery text (on display in May 2018) confirms that Klenze manipulated the temple's relative positioning for aesthetic reasons.

cannot be relied upon to correct or regulate testimonial practice. Recordings could become no more authoritative than paintings, only as reliable as the reputation of their creator. A recording is then just another node in the testimonial web, equally subject to partisan repudiation or inflation. And there is less reason to be responsible in one's testimonial practices if recordings offer no more than a source of conflicting testimony.

Cavedon-Taylor and Hopkins make similar points about the effects of Photoshop on the epistemology of still photographs.<sup>33</sup> Indeed, this worry was raised by Barbara Savedoff as early as the year 2000, 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."<sup>34</sup>

But this returns us to the challenge with which I began this section. If recordings have a perceptual-type epistemic role similar to still photographs, and if the reliability of photographs has been under assault from Photoshop for two decades, why haven't we *already* seen the sort of epistemic catastrophe I described in the last section? It seems like we've adapted well enough to widespread knowledge that still photos can be manipulated – so why won't we adapt just as well to deepfakes?

### **The distinctive threat of deepfakes**

---

<sup>33</sup> Hopkins 2012, 723. Cavedon-Taylor actually goes further: if our awareness of digital manipulation thoroughly erodes our trust in photos, then we might conclude that "photographically based belief, in order to be rationally grounded, must be formed on the basis of positive reasons for thinking the photograph has been reliably produced. This will have the effect of rendering photographically based knowledge exclusively inferential in nature, and not testimonial" (Cavedon-Taylor 2013, 296).

<sup>34</sup> Savedoff 2000, 202.

In fact, we've known that photographs are vulnerable to manipulation since long before Photoshop. In 1920, Arthur Conan Doyle published an article vouching for a series of photographs of fairies taken by two young girls in West Yorkshire. Of course, the creator of Sherlock Holmes would not have done so without ruling out alternative explanations. First he sent the negatives for inspection by Kodak technicians, who refused to provide a certificate of authenticity, though they agreed the materials "show no signs of being faked".<sup>35</sup> Conan Doyle then considered that the girls might have simply drawn the fairies on cardboard and posed them very carefully. But, Conan Doyle decided, "the girl's own frank nature is, I understand, a sufficient guarantee for those who know her. Mr. Gardner, however, tested her powers of drawing, and found that, while she could do landscapes cleverly, the fairy figures which she had attempted, in imitation of those she had seen, were entirely uninspired, and bore no possible resemblance to those in the photograph."<sup>36</sup> (Sixty years later, the women admitted this was exactly what they had done – though they insisted there really had been fairies around.)

Infamously, Stalin employed whole teams to bring scalpel and airbrush down upon the photographic records of disfavored comrades.<sup>37</sup> In one extraordinary case, the original 1926 image poses Stalin alongside three others. Over later years a succession of cropped and retouched versions of the photo shed comrades as decisively as the Politburo. The last image – Stalin all alone – has been so tortured that it is plainly more painting than photograph.<sup>38</sup>

So we have known for decades that photos can be manipulated, yet our testimonial norms seem largely intact. Why, then, should we worry much about deepfakes? There are at least four reasons to worry.

---

<sup>35</sup> Paul Smith (1997)

<sup>36</sup> Arthur Conan Doyle (1920)

<sup>37</sup> See David King (1997) *The Commissar Vanishes: The Falsification of Photographs and Art in Stalin's Russia*. New York: Henry Holt and Company.

<sup>38</sup> These images are in the public domain and may be viewed at [https://commons.wikimedia.org/wiki/File:Soviet\\_censorship\\_with\\_Stalin2.jpg](https://commons.wikimedia.org/wiki/File:Soviet_censorship_with_Stalin2.jpg) [[accessed March 11, 2019]]

First, there is a psychological difference between still photographs and audio-video recordings. Recordings are typically more gripping, perhaps because they extend over time and support articulated narratives. People happily spend money to gather in theaters and watch movies, but they don't (usually) do so to view still photographs. Perhaps recordings simply *feel* more like reality than photographs. We rarely experience the world in static images, but a well-executed recording mimics our everyday perceptual activities. This suggests we may be more psychologically invested in the reality of recordings than still photographs. When backstop crises force us to confront the growing unreliability of recordings, the skeptical consequences may be amplified by their psychological power.

Second, deepfakes, unlike familiar Photoshop images, are generated by a machine learning process that permits *efficient response* to epistemic challenges. Currently, if we doubt someone's testimony about an event, we can challenge them to produce a recording. And if we doubt the first recording they provide, their ability to produce a second corroborating recording would be strong evidence in their favor. For example, in 2016 Trump campaign manager Corey Lewandowski was accused of grabbing the arm of a reporter and pulling her away from questioning the candidate. The campaign quickly dismissed the charge, claiming that "not a single camera or reporter of more than 100 in attendance captured the alleged incident".<sup>39</sup> Reporters took up the challenge, quickly locating two different video recordings, from alternate angles, showing Lewandowski doing just what was alleged.

Part of why the campaign's challenge was compelling – and why successfully meeting the challenge was even more so – is that we currently do not expect anyone to be able to swiftly produce fake video recordings tailored to novel (purported) facts, in the way that we know Photoshop is relatively cheap and easy to use. Though Hollywood can produce extremely compelling videos of unreal

---

<sup>39</sup> Reena Flores (2016), 'Donald Trump's campaign denies getting rough with reporter'. *CBS News* March 10, 2016. <https://www.cbsnews.com/news/trump-campaign-responds-to-charges-of-getting-physical-with-reporter/> [[accessed March 11, 2019]]

things, we know that this takes enormous resources of time, money, and human talent. That's why it is implausible to explain away effectively-met challenges to recordings; currently, the idea of overnight faked recordings sounds more like a conspiracy theory than a plausible debunking. But this thought will become less powerful after we have witnessed a few backstop crises. Once deepfake technology is widely known, it will be possible for public figures in situations like Lewandowski's to simply insist that any purported recording evidence is just more fakery – and who will be able to say they are wrong?

Third, the efficiency of machine learning will enable mass production of deepfakes at a similar rate to textual fake news. This will allow epistemic malefactors to 'spam' the epistemic environment with conflicting false videos of prominent events, in much the way they currently do with fake news. This is an established tactic of authoritarians seeking to disrupt public information channels. Neutralizing effective public epistemology doesn't require tricking people into believing any particular falsehood. It just requires convincing them not to trust *any* information source. Once information channels are thoroughly saturated with obvious garbage, people fall into a sort of epistemic learned helplessness, where they give up trying to critically assess information and simply believe whatever conforms to their worldview.<sup>40</sup> There is every reason to expect such malefactors will use deepfakes in the same way.

Once that happens, we will see the vulnerability of recordings demonstrated frequently and shockingly, in a way that hasn't (yet) happened to still photographs. Of course, it's possible that photos will also soon be undermined by machine learning mass-production. My point is only that the machine learning technology behind deepfakes leads to a predictable threat distinct from earlier photo manipulation techniques.

---

<sup>40</sup> (Tufekci 2017) ;Lynch, Michael. (2016, November 28). Fake News and the Internet Shell Game. *New York Times*. Retrieved from <https://www.nytimes.com/2016/11/28/opinion/fake-news-and-the-internet-shell-game.html>

For a worked example of how the Russian regime has used this technique in the past, see Snyder (2018, chapter 5).

Finally, and perhaps most importantly, the difference between Photoshop and deepfakes may be simply this: the epistemic authority of still photographs *has* already been eroded by Photoshop, but our evidential practices around photographs are, like verbal testimony, currently still undergirded by the epistemic backstop of recordings. Perhaps the evidential status of still photos has *already* shifted from vehicles of perceptual knowledge to mere testimonial knowledge. But so long as recordings still serve their backstop functions, they can acutely correct and passively regulate photographic evidence in the same way as testimonial practice. We already tend to trust video recordings over still photographs when they appear to conflict. A still photo of a single moment can mislead about causal relations among objects in a way that is often quickly resolved by a corresponding video.

But if this is true, then we already implicitly place more epistemic weight on recordings than we appreciate. The epistemic backstop regulates not only our testimonial practices, but also our use of others sorts of documentary evidence. This suggest that the threat deepfakes pose to the backstop may be even more consequential, exposing neglected decay in our epistemic relations to photos along with undiagnosed vulnerability in testimony.

### **Toward a worrisome epistemic future**

Where does this leave us? If my worries are right, then in the near future we face sudden collapse of the backstop to our testimonial practice, as we can no longer trust that recordings provide authoritative correction. Following highly public backstop crises, we may find that our adherence to testimonial norms of competence and sincerity becomes progressively less reliable, as people realize there is rarely an independent check on their testimony. What happens then?

One answer is surprisingly positive. You might think that we never *should* have put so much trust in recordings. After all, recordings have never been strictly invulnerable from manipulation; given

extensive time and resources, skilled fakers could produce convincing videos long before the rise of machine learning. So, it might be a good thing if deepfakes shake us free from our unreasonable reliance on recordings.<sup>41</sup>

I think this is the wrong attitude to take. It assumes an implausibly Spartan epistemology, one where we restrict our evidence to near-infallible technologies. I think this is an unrealistic epistemic standard that would deprive us of valuable sources of information quite often. To put it another way, I think that our epistemic reliance on recordings *has been* reasonable, even if it will soon cease being so.

But you don't have to agree with me about this in order to see the danger of deepfakes. For, even if abandoning our reliance on recordings might be epistemically admirable, there will serious *transition costs*, and these ought to worry everyone.

For better or worse, we *have* developed a web of epistemic norms assuming reliance upon recordings. In the developed world, there is no one living today who remembers an epistemic environment preceding that reliance. Video and audio recordings, in existence longer than any of us, have always structured our lives. To really appreciate what their discredit would mean for our testimonial practice, we must look back to a time when people relied only on the testimony of eyewitnesses and newspapers. If I am right about the effects of deepfakes, we may be very suddenly plunged into just such an environment, without preparation or training.

There were, of course, testimonial norms before the 19<sup>th</sup> century creation of recording technology. Those norms might be adequate if we could return to them. But how do we get back there? None of us have the benefit of life-long training in an unregulated testimonial environment. We may become dangerously credulous, or perhaps reactively paranoid. I cannot predict the future, but I am

---

<sup>41</sup> I thank audience members at the 2019 Central APA in Denver for convincing me of the merit of this objection.

confident that we will not quickly rediscover 19<sup>th</sup> century testimonial norms without a lot of trouble along the way.<sup>42</sup>

More fundamentally, I worry that 19<sup>th</sup> century testimonial norms simply won't *work* in a modern world of instant global communication. We may have been fortunate to enjoy the testimony-regulating effects of the epistemic backstop during the last 150 years of rapid technological and social change. But we may now be forced to navigate future changes without that protection. In a social epistemic environment already plagued by fake news malefactors and authoritarian deceivers, I am less than fully confident we will weather the transition with our social and political systems intact.<sup>43</sup>

---

<sup>42</sup> Among other problems, we should worry about losing the epistemically equalizing benefits of recordings. 19<sup>th</sup> century epistemology left members of marginalized groups especially vulnerable to testimonial injustice (Fricker 2007). Reliable video recordings have sometimes provided marginalized groups with the means to confront disbelieving publics, as with recent smartphone videos of police brutality against African Americans. (I owe this important point to Daniel Saunders.)

<sup>43</sup> I would like to thank Daniel Saunders for invaluable research assistance on this project. I also owe thanks to Leah Cohen for assistance on an earlier project – and for bringing deepfakes to my attention. Thanks also to audiences at the 2018 'Fake Knowledge' conference at the University of Cologne, the 2019 Central APA in Denver, the 2019 'Ignorance in the Age of Information' conference at Scripps College, and the University of St. Andrews. This research was supported by the VISTA initiative at York University.

---

## References

- Baier, Annette 1986, 'Trust and antitrust,' in *Ethics* Vol. 96 No. 2
- Benjamin, Walter 1936[1969], 'The Work of Art in the Age of Mechanical Reproduction'. in *Illuminations*. Ed. H. Arendt. (New York: Schocken) 217–251.
- Cavedon-Taylor, Dan 2013, 'Photographically Based Knowledge,' in *Episteme* Vol. 10, No. 3
- Chesney, Robert and Danielle Keats Citron Forthcoming 2019, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, in *California Law Review* <https://ssrn.com/abstract=3213954>
- Coady, C. A. J. 1992, *Testimony: A Philosophical Study* (Oxford: Oxford University Press)
- Cohen, Jonathan and Aaron Meskin 2004, 'On the Epistemic Value of Photographs,' In *The Journal of Aesthetics and Art Criticism*, Vol. 62, No. 2
- Conan Doyle, Arthur 1920, 'Fairies Photographed', reprinted in the Arthur Conan Doyle Encyclopedia at [https://www.arthur-conan-doyle.com/index.php?title=Fairies\\_Photos](https://www.arthur-conan-doyle.com/index.php?title=Fairies_Photos)
- Fricker, Miranda 2007, *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press).
- Goldberg, Sanford 2010, *Relying on Others: An Essay in Epistemology* (Oxford: Oxford University Press)
- Hopkins, Robert 2012, 'Factive Pictorial Experience: What's Special About Photographs?' in *Noûs* Vol. 46, No. 6
- Jin et al 2017, 'VoCo: Text-based Insertion and Replacement in Audio Narration', in *ACM Transactions on Graphics* 36(4):96. <https://doi.org/10.1145/3072959.3073702>
- Jones, Karen 1996, 'Trust as an affective attitude,' in *Ethics* Vol. 107 No. 1
- King, David 1997, *The Commissar Vanishes: The Falsification of Photographs and Art in Stalin's Russia* (New York: Henry Holt and Company)
- Lackey, Jennifer 2007, 'Norms of assertion,' in *Nous* Vol. 41 No. 1
- Lackey, Jennifer 2008, *Learning From Words: Testimony as a Source of Knowledge* (Oxford: Oxford University Press)
- Liu et al. 2018, 'Neural Animation and Reenactment of Human Actor Videos', in *arXiv:1809.02658v1* <https://arxiv.org/pdf/1809.03658.pdf>
- Quiring, Erwin, Daniel Arp and Konrad Rieck 2017, 'Fraternal Twins: Unifying Attacks on Machine Learning and Digital Watermarking', in *arXiv: 1703.05561v1*
- Savedoff, Barbara 2000, *Transforming images: How photography complicates the picture* (Ithaca, NY: Cornell University Press)
- Smith, Paul 1997, 'The Cottingley Fairies: The End of a Legend', In Peter Narvez (ed.), *The Good People: New Fairylore Essays* (Lexington: University of Kentucky Press: 371-405: 384).
- Snyder, Timothy 2018, *The Road To Unfreedom: Russia, Europe and America* (New York: Tim Duggan Books)

Tufekci, Zeynep 2017, *Twitter and Tear Gas: The Power and Fragility of Networked Protest* (New Haven and London: Yale University Press)

Vosoughi, Sorouah, Deb Roy and Sinan Aral 2018, 'The spread of true and false news online', in *Science* Vol. 359, Issue 6380

Walton, Kendall 1984, 'Transparent Pictures: On the Nature of Photographic Realism,' In *Noûs* Vol. 18, No. 1

Williamson, Timothy 1996, 'Knowing and asserting,' in *Philosophical Review* Vol. 105 No. 4

Woodward, Bob and Carl Bernstein 1976, *The Final Days* (New York: Simon & Schuster)