Synthese (2023) 201:30
https://doi.org/10.1007/s11229-022-04012-2

ORIGINAL RESEARCH

Designed to abuse? Deepfakes and the non-consensual diffusion of intimate images

Marco Viola[1] · **Cristina Voto**[2]

Received: 25 July 2022 / Accepted: 14 December 2022 / Published online: 13 January 2023

Abstract
The illicit diffusion of intimate photographs or videos intended for private use is a troubling phenomenon known as the diffusion of Non-Consensual Intimate Images (NCII). Recently, it has been feared that the spread of deepfake technology, which allows users to fabricate fake intimate images or videos that are indistinguishable from genuine ones, may dramatically extend the scope of NCII. In the present essay, we counter this pessimistic view, arguing for *qualified* optimism instead. We hypothesize that the growing diffusion of deepfakes will end up disrupting the status that makes our visual experience of photographic images and videos epistemically and affectively special; and that once divested of this status, NCII will lose much of their allure in the eye of the perpetrators, probably resulting in diminished diffusion. We conclude by offering some caveats and drawing some implications to better understand, and ultimately better counter, this phenomenon.

Keywords Social epistemology · Deepfake · Visual culture · Photography · Testimony

1 Introduction

One morning, while checking the mail inbox, one of us stumbled upon a text from a self-proclaimed hacker. "The previous time you went to the porno online sites my spyware ended up being activated inside your computer system which ended up recording an eye-catching video footage of your masturbation play by triggering your webcam".

✉ Marco Viola
marco.viola@uniroma3.it

Cristina Voto
cristina.voto@unito.it

1 Department of Philosophy, Communication, and Performing Arts, Roma Tre University, Rome, Italy

2 Department of Philosophy and Educational Sciences, University of Turin, Turin, Italy

Then, the hacker showed they knew an (old) password of ours. The remainder of the mail followed a predictable plot, threatening to share the video-footage with some random contacts unless a 2000 BitCoin donation were made within the following 24 h.

A quick online query confirmed the impression that this was a crazy attempt at "sextortion". The very same mail had been sent to thousands of recipients, whose mail addresses and old passwords had been retrieved from a data breach on some abandoned website. Of course, failure to make the donation did not result in the spread of any video footage. The web is replete with millions of scams like this, and the threat of displaying intimate images is one of the most popular strategies for blackmailing someone. Yet, despite rationally knowing that no hacker could possibly have access to such nasty materials, it took a while to shrug off the sense of uneasiness accompanying the threat.

Alas, there are cases in which the sensation of being threatened cannot be shrugged off that easily. These are the cases in which intimate images actually get shared without the consent of the portrayed victim. The provocative movie *Bad Luck Banging or Loony Porn* (2021) by Radu Jude ironically recounts the unfortunate consequences faced by a history teacher, Emi Cilibiu, after a private sex videoclip showing her having intercourse with her husband goes viral. The psychological and social hardships Emi undergoes culminate in a sort of popular (and populist) trial in which her students' parents vote for or against firing her from the school.

Other than providing a vivid exemplification of the phenomenon of non-consensual diffusion of intimate images (enacted or threatened), the two examples above provide us with grounds for introducing the question we are going to address: what if intimate images no longer needed to be captured by a camera in a given place at a given time because an algorithm existed that could fabricate them? This question is not about a distant future. Rather, it refers to the so-called deepfake technologies, which are becoming increasingly sophisticated and increasingly available. With such a powerful tool at hand, the abovementioned hacker would not even have to pretend they had compromising video footage: they could fake it. If so, everybody would be permanently under threat. Or not?

The concerns expressed by journalists and scholars dealing with such issues (e.g. Cole, 2017; Chesney & Citron, 2019; Gosse & Burkell, 2020; Flynn et al., 2021) offer grounds for fearing that the rapid development of deepfakes has already begun making these abuses ubiquitous, exponentially increasing their potential for harm. Yet, somewhat counterintuitively, we argue that this (forthcoming) ubiquity provides some cause for prudent and *qualified* optimism. In a nutshell, we claim that, by eroding the special epistemic and affective status of photographic images and videos, the diffusion of deepfakes may end up weakening the voyeuristic allure of non-consensual intimate images in the eyes of the perpetrators who spread them, as well as their grip on the victims. Our optimistic conclusions do not warrant in any way a lax attitude toward the non-consensual diffusion of intimate images. Quite the contrary: our aim is to provide a better understanding of the phenomenon and possibly to predict some of its possible evolutions, so as to better counter it.

Our discussion is structured as follows. In the next Sect. 2, we introduce the terminology and offer a sociological overview of the phenomena at hand. Then (in Sect. 3),

after having reflected on the impact that deepfakes may exert over this landscape, we flesh out our argument. In essence, the attraction of non-consensual intimate images for perpetrators and their potential to harm the victims hinge upon the current epistemic and affective status of photographic images and videos (Sect. 4). However, the diffusion of deepfakes will subvert this status, provided that people become aware of their potentialities (Sect. 5). We conclude by pointing out some caveats and by flagging possible problems that may require a vigilant eye in the future (Sect. 6).

2 Framing the problem: non-consensual diffusion of intimate images and other image-based sexual abuses

Unfortunately, the abovementioned *Bad Luck Banging or Loony Porn* does not belong to the sci-fi genre: newspapers are full of similar stories. Such cases are often labeled "revenge porn" in the popular press, but there are good reasons to reject this label.

The term "revenge porn" initially designated the practice of non-consensually sharing images or videos of former sexual partners, often with the aim of humiliating them (Hearn & Hall, 2019). However, its usage was soon extended to refer to all non-consensual sharing of intimate contents. Scholarly research shows that "revenge" over ex-partners is but one of the many reasons behind the non-consensual diffusion of intimate images or videos: scholars who interviewed perpetrators or analyzed their discourses (e.g. Semenzin & Bainotti, 2020; Henry et al., 2020; Naezer & van Oosterhout, 2021) found other (non-mutually exclusive) reasons, e.g. obtaining sexual gratification, pulling "pranks", or even reinforcing bonds within online communities hegemonized by toxic masculinity[1].

It has also been noted that speaking of "revenge" insidiously promotes victim-blaming, as it suggests that the victim has done some harm that deserves a punishment. For these and other similar reasons, scholars and civil rights activists are considering the pros and cons of different labels to refer to these phenomena (see Maddocks, 2018).

The term 'non-consensual pornography' gained some traction, especially in the US. But it has also met some criticisms. For instance, a scholar interviewed by Maddocks (2018, p. 5) deems 'pornography' inappropriate because "These images were not created for public consumption". Indeed, many NCII were not even meant to be a sexual performance in the eye of the portrayed victim; it is the perpetrators' and consumers' gazes that sexualize them.

Given the prominence of the *male gaze* in our society, female bodies are particularly vulnerable to sexualization. As we learn from film studies (notably from Mulvey, 1975), the massive spread of mainstream cinema has promoted and buttressed

[1] The term "toxic masculinity" usually refers to that set of harmful cultural rules about masculinity that group together a set of stereotypes concerning the domination of men in society, leading to misogynistic and homophobic drifts. However, similarly to "revenge porn", the label "toxic masculinity" is a conceptually thorny one: for instance, it promotes the idea that misogyny and sexism are problems that afflict only a certain number of toxic and therefore sick men; it underestimates the fact that phenomena such as misogyny are rooted in culture, not in biology (Harrington, 2021). A more complex perspective permeates these pages. The aporias implicit in the terminology are often known. Nevertheless, the syntagma is widely used in contemporary debates because it succeeds in conveying with compelling force the possibility that non-hegemonic forms of masculinity emerge and reproduce.

a habit of conceiving the (idealized) female body as an object of pleasure available to the sexualizing gazes of the onlookers (the to-be-looked-at-ness as described by Mulvey). Taking into account the social normativity of gazing anticipates something that we will see in greater detail later on, namely, that the allure and the harmfulness of images do not depend merely on their visual properties (or audio-visual properties in the case of videos), but also on the social context in which they circulate.

Following Henry et al. (2020), in this essay we will adopt the wider expression 'diffusion of Non-Consensual Intimate Images' (NCII), which includes both photographs and videos. To emphasize the continuum with other forms of abusive behaviors, some scholars have placed NCII within the broader category of image-based sexual abuse (IBSA; McGlynn et al., 2017; Herny et al., 2020: ch. 1). This broader label includes cases in which the non-consensual sharing of images is not (or not only) enacted, but also threatened – as in the sextortion attempt reported at the beginning of this paper. Ascribing NCII to the broader category of sexual abuse highlights some common features they share with other abusive behaviors: for instance, the fact that they have a stronger effect on women and on some minorities; and that they often occur in online communities held together by toxic masculinity bonds (see notably Semenzin & Bainotti, 2020). Finally, highlighting their abusive nature does justice to the sense of violation and the permanent sense of threat felt by the victims. Given how hard it is to permanently delete certain content once it hits the Internet, some of the victims interviewed by Henry et al. (2020) manifested "a sense of ongoing, existential threat which can cast a shadow over [their] lives" (p. 58), because (quoting a victim's interview) "the images could be re-shared, or re-emerge online, [and] new people could see these intimate images" (p. 57).

On the one hand, NCII and other forms of IBSA are not a new phenomenon. Indeed, the recent drama series *Pam & Tommy* recounts the theft and leakage in 1995 of a private sex tape featuring Pamela Anderson and Tommy Lee, and the harm it caused—especially to her—once it spread virally. However, older cases can also be invoked: for instance, in 1953 the first issue of the magazine *Playboy* contained naked photographs of the famous actress Marylin Monroe. These pictures were taken before she reached fame as an actress, when she posed as a pin-up for a local calendar. But the payment received and the consent given by her only pertained to a few hundred copies of that local calendar, not a magazine distributing thousands of copies across the US several years later.

On the other hand, technological development has greatly facilitated taking and spreading pictures and videos, thus promoting a quantum leap in the magnitude of NCII and other IBSA. Already at the beginning of the Millennium, legal experts alarmed by the spread of the non-consensual diffusion of intimate images were concerned that "the technology necessary to secretly capture images on videotape today is both inexpensive to purchase and relatively easy to operate" (Calvert & Brown, 2000, p. 480). The diffusion of smartphones, which allow users to take and share intimate images in the blink of an eye, has further facilitated the production and/or spread of

non-consensual sexualized images, especially in recent times, due to the spread of *sexting*, i.e. digitally-mediated sharing of intimate contents.[2]

It is nearly impossible to make an accurate estimate of how widespread the problem of NCII is due to several issues: for instance, given also that laws to sanction them have been recently promoted in several countries (e.g. Caletti, 2021; Jochelson et al., 2021), perpetrators may be reluctant to confess their misdeeds. Nevertheless, existing data show that the problem is widespread. A recent large-scale survey with semi-structured interviews, encompassing 6109 respondents from 16 to 64 years of age across Australia, New Zealand, and the United Kingdom, found that 1 in 3 respondents have been victims of some sort of IBSA, and 1 in 6 have perpetrated at least one (Henry et al., 2020).

However, another technological quantum leap is forthcoming, which may sustain a further amplification of the plague of IBSA, namely, the rapid development and increasing availability of deepfake software allowing for large-scale home-made fabrication of NCII.

3 Enter deepfakes

The term "Deepfake" indicates artificial images or videos that resemble actual photographs or videotapes. They often exploit the power of deep neural networks (hence "deep")[3], and are often produced with the intent to deceive viewers into believing that the fabricated content is real (hence "fake"). To be sure, nothing prevents the creation of deepfakes that are *overtly* fake. For instance, people can have fun swapping their face onto some actor's body in a topical movie scene. Despite being illegal in many countries and being banned from many porn platforms, the web is still replete with deepfake videos depicting celebrities' faces placed atop naked bodies that are obviously not their own, engaging in sexual activities that the viewers clearly know celebrities not to have performed. While these kinds of overt sexual deepfakes can also harm by reinforcing the male gaze (see sect. 6), in the remainder of this essay our main focus are the *covert* ones, i.e. those that have the potential to deceive.

Just as IBSA pre-dates the digital age, counterfeit images are far older than deepfakes—possibly as old as photography itself. For instance, in 1869, the photographer William H. Mumler faced a trial for fraud related to his spirit photographs, allegedly depicting ghosts of dead persons (Fineman, 2012, pp. 22–24). Yet, fabricating such

[2] Texting is particularly widespread among adolescents. A recent study reports that 40.9% of a sample of Belgian adolescents (n = 549, aged 12–18) engaged in at least some form of sexting during the lockdown (Maes & Vandenbosch, 2022).

[3] A powerful technology for producing deepfakes involves the competing activity of two Neural Networks, a generator and a discriminator. The technology, known as Generative Adversarial Networks, is depicted by its developers as "a discriminative model that learns to determine whether a sample is from the model distribution or the data distribution. The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency" (Goodfellow et al., 2014: 1). The fundamental feature for the achievement of the goal is to conceal the complexity of the manipulation practice behind the apparent highlighting of a referent.

images required skillful methods, ranging from "relatively simple techniques of double exposure to more elaborate trickery involving a microscope lens or a glass plate containing a tiny image of the spirit 'extra', inserted into the camera before the picture were taken" (p. 24). In contrast, deepfake technologies are making image manipulations widely accessible and quite effortless. An example is the controversial app DeepNude, which takes a picture of a clothed body as input, estimates its morphological features, e.g. body size and skin color, and uses them to provide a naked female body as output. This app prompted a huge controversy and was withdrawn by its developer after only 3 days. Alas, as we mentioned, it is hard to permanently delete something once it reaches the web, hence certain versions of DeepNude and similar software still circulate illegally. Moreover, even the above-mentioned faceswap apps can be easily exploited to create deepfake pictures or videos.

Thus, while until recently IBSA required the actual acquisition of someone's photograph or video, with the progressive refinement and diffusion of deepfake technologies this constraint may be eased—and perhaps even *erased*. Anyone can become a potential target anytime. If a photograph of a clothed person suffices to fabricate a realistic naked counterpart, abstaining from sexting or from leaving digital traces of one's own intimacy may become pointless[4]. A safe strategy would be to have no images of oneself ever captured by digital devices. But in modern digital societies, that seems hard to achieve for many people.

Recently, Flynn et al. (2021) provided evidence that "deepfake and digitally altered imagery is an emergent form of abuse that has the potential to generate significant harm, with some populations likely to be more vulnerable to experiencing it" (p. 14). They re-examined the data from the abovementioned multi-country survey (Henry et al., 2020), this time focusing specifically on IBSA mediated by deepfakes or other forms of image alterations. 7.6% of their sample (n = 466) had engaged in at least one form of IBSA involving creating, sharing, or threatening to share manipulated contents; and 14% (n = 864) had been victims of at least one form of abuse. Prospectively, they noted, "as the simplification and reach of [deepfake and other digital alteration technologies] increases, the risk of people experiencing harm also increases" (p. 2).

Are we heading toward a future in which we are all potentially vulnerable to the "existential threat" brought about by IBSA? As anticipated, while we must not lower our guard in relation to IBSA, we see some grounds for qualified optimism. Our argument runs as follows:

> [Premise 1] The allure of deepfake NCII, as well as their potential to harm, heavily relies on the special epistemic and affective status that we currently associate with photographic images and videos. But
>
> [Premise 2] the increased (awareness of) spread of deepfakes will progressively erode their special epistemic status and possibly their affective status.
>
> [Conclusion] In the long run, the very diffusion of deepfake NCII will downplay their allure and their potential to harm.

[4] Some scholars have already expressed skepticism about strategies to contrast NCII based on abstinence, e.g. discouraging people (and especially female adolescents) from texting, as they may run the risk of reenforcing the script of rape culture by shifting the blame from the perpetrators to the victims (Powell & Henry, 2014; Naezer & van Oosterhout, 2021).

We defend premises 1 and 2 in the next sections (Sect. 4–5), whereas in the last section we address some possible objections and caveats (Sect. 6).

4 The status of photographic images and videos and its role in NCII and IBSA

Philosophers have long debated the status of photographic images. Several accounts have sought to explain the peculiarities of our visual experience of photos (or what we take to be photos) and in what respects this differs from other visual experiences: on the one hand, the visual experience of other kinds of images (e.g. hand-made drawings); on the other hand, that of direct vision.

Most of these discussions begin with Walton's (1984) controversial thesis that photographic images are *transparent.* In his view, while on the one hand photographs are akin to other kinds of pictures (like paintings), in that they enable fictional vision of their depicta, on the other hand they are at the meantime also tools for vision (like mirrors or telescopes), i.e. they enable actual if indirect seeing of their depicta. Accordingly, while looking a painting of one's dead relatives only brings about *imagining* seeing them, "we see, quite literally, our dead relatives themselves when we look at photographs of them" (Walton, 1984, p. 252). While it soon became clear that the notion of seeing advocated by Walton was revisionist, his controversial thesis proved quite resilient, as he has thrown down the gauntlet for the epistemology of photographs, i.e. accounting for the special epistemic status of photographs (see Costello & Philips, 2009). How to take up the gauntlet without committing to a revisionist notion of 'seeing'?

According to Currie (1999), photographs are epistemically distinct from other kinds of images because they are 'traces', i.e. images whose content is counterfactually dependent on the photographed object, in a manner that is not mediated by beliefs. Cohen & Meskin, (2004; see also Meskin & Cohen, 2010) focus on the visual information conveyed by photographs, stressing that (like direct vision) it is usually richer than the information provided by drawings, but (unlike direct vision or vision mediated by tools like mirrors) fails to deliver information about *egocentric* visual properties. While these theories mainly hinge on some ontological properties of photographs, other scholars place more emphasis on the onlooker. Cavedon-Taylor (2013, 2015) shifts the focus onto cognitive phenomenology, showcasing how the etiology of the beliefs we acquire via pictorial experience of photography is usually immediate rather than inferential. Whereas Hopkins (2012) holds that the key property of photographs is that they yield factive pictorial experience, i.e. what we see in the photo typically corresponds to what has been photographed. This factiveness is grounded not only in the truth-preserving process through which photographic images are made, but also in the widely shared assumption that the steps necessary to develop photographs from negatives used to be influenced by norms that ensured truth-preserving processing, which are largely maintained also in the production of digital photographs.

Now, while differing in the nuances of their explanations of *why* photographic images have a special status, these authors are consistent with respect to *what* this status entails. First, all concur that photographic pictures have a privileged *epistemic*

status with respect to other kinds of images, e.g. handmade drawings. Walton (1984) puts it in terms of "realism". Meskin & Cohen (2010, p. 70) note that "we are inclined to trust them in a way that we are not inclined to trust even the most accurate of drawings and paintings". Hopkins (2012, p. 710) stresses that "unlike our experience of other pictures, our experience of photographs is factive: it is guaranteed to reflect the facts". Borrowing Peirce's terminology: photographic images are taken to be *seals of indexicality*[5].

Following Cavedon-Taylor, we deem it useful to compare our epistemic attitude toward photographic images with that which we have toward testimonial sources. A dichotomy lies at the foundation of social epistemology, namely, the contrast between non-reductive and reductive accounts of testimonial knowledge (for an introductory exposition, see Leonard, 2021, Sect. 1). Non-reductionists, championed by Thomas Reid, hold that whenever a testimony t claims some proposition p, barring positive reasons for believing that t is untrustworthy, we are rationally entitled to believe p. In a nutshell, they are *trusting toward other agents by default*, and remain trusting until proven wrong. On the contrary, reductionists are *skeptics by default*: following the lead of David Hume, they maintain that we are entitled to believe p only if we have positive reasons for believing that t is trustworthy. Now, according to Cavedon-Taylor, we are naturally biased toward non-reductionism: "we assent to the content of our pictorial experiences before photographs by default; that is, so long as we do not possess reasons for thinking the photograph *un*creditworthy" (2015, p. 77)[6]. We agree with him that this is probably the default doxastic attitude for most people. Moreover, we believe that, while interacting with others, most people also expect this assumption to be their default attitude. This expectation about the attitude of others can also be found in the developer of DeepNude. In the free version of the app, the pictures of undressed subjects generated as output were partially covered by a semi-transparent watermark signaling that they were "fake nudes". Were the programmer convinced that our doxastic attitude falls closer to Hume's dictum than to Reid's, this watermark would be unnecessary, as the burden of proof would fall on the shoulders of those who want to argue that the picture depicts some genuine fact.

Beyond their privileged epistemic status, photographs seem to enjoy a special status vis-à-vis other images with respect to the affective reaction they elicit. Walton (1984) and Hopkins (2012) speak of 'intimacy'; Currie (1999), Walden (2016) and Anscomb

[5] In Peirce's well-known account, signs are sorted into a tripartition based on which relation links them with their referent: *Icons* signify by virtue of a resemblance with respect to some quality; *Symbols* by virtue of some conventional connection; whereas *Indexes* by virtue of some causal dependence (Peirce, 1894/1982). Notably, the three varieties of signs are not mutually exclusive. Photographs (and films) are a prime example, as they are sometimes said to be *iconic indexes*: not only they represent in virtue of some resemblance, as portraits or sculptures, but this resemblance is rooted in some tight causal connection between the photographic (filmic) images and what they represent (see Friday, 2002; Sadowski, 2011). In what follows, we will stress indexicality rather than on iconicity because this is what marks out photographs from other kinds of pictures like portraits or deepfakes.

[6] Notice that while the debate between non-reductionists and reductionists in the epistemology of testimony is often cast in normative terms ("should we trust a testimony?"), here we align ourselves with Cavedon-Taylor in considering mainly the descriptive-cognitive facet of the epistemology of photographic pictures ("do we usually trust photographs?"). However, the distinction is not always neat. For instance, authors that focus on the properties and the etiology of photographs (e.g. Currie, 1999; Hopkins, 2012) may have more normative aims in mind.

(2022) of 'contact'; Petterson (2011) of 'proximity'. These different words express the same feeling. Relevant to our purpose here, a comparison that some authors invoke to underscore their point is that naked or pornographic photographs are usually more arousing than realistic hand-drawings (Walton 1984, Cavedon-Taylor 2015).

We maintain that these features of photography, i.e. their privileged epistemic and affective status, are deeply entrenched[7]. Consider the example of a photograph allegedly depicting some war horror: we have a hard time imagining that the affective reaction of someone who maintains that the picture comes from the set of a war movie can be as intense as that of someone who thinks that the picture comes from an actual battlefield. Supporting this intuition, some psychological studies conducted during the Seventies report that watching videos depicting violence increases angered subjects' aggressiveness and arousal if the videos are presented as representing actual violence, but not if they are presented as staged violence (Thomas & Tell, 1974; Geen, 1975).

Despite several differences between photographic images and videos, the above-mentioned epistemic and affective properties of photographic images (i.e. being taken as a seal of indexicality, eliciting a sense of intimacy) are also shared by videos to an interesting extent. We think that these properties underlie the allure of NCII in the eye of the perpetrators and the power of these images (and videos) to harm the victims of IBSA.

Before DeepNude was withdrawn, people might spend $50 to get the premium version, which worked exactly like the free version but for one feature: the images it yielded as output no longer had the invasive watermark signaling that the image was a fake[8]. What people paid for, thus, was not the access to some visual contents: all the relevant aesthetic properties of the image were already in place in the free version. Instead, they paid to have the "effect of reality deactivator" removed. By getting rid of this skepticism-activating watermark, they had less trouble pretending that the image was obtained by means of an actual photo shoot; and it could be circulated professing to be a real photograph, rather than a deepfake.

In their attempt to understand the motivation underlying the consumption and diffusion of NCII based on the scant and indirect evidence currently available, Henry et al. (2020, ch. 5) trace a parallel with amateur pornography. As noted by some scholars investigating digital media (e.g. Paasonen 2010; Byron et al. 2021), a significant number of porn consumers prefer porn material that is (or pretends to be) amateur over the professional contents produced by the porn industry. This is somehow puzzling: why should a self-made, shaky video with poor illumination and taken from a bad angle sometimes be preferred to a high-budget quality product with professional illumination and several camera operators? The main reason, according to these authors, is that this "person-next-door" halo triggers a sense of realism; and that this perceived authenticity, in turn, heightens the viewers' engagement[9]. This might explain the proliferation

[7] Walden (2016) and Anscomb (2022) have made quite a strong case that the epistemic and the affective attitude toward pictures can come apart. However, in the next session we shall explain why we think this does not impinge our argument.

[8] In the premium version of DeepNude, a small watermark signaling the fake origin of the image was still present, although it could be cropped with ease even with the simplest graphic editing software.

[9] Cf. the pioneering use of blurred recordings in the horror movie *The Blair Witch Project* to imply that it could have been real video footage.

of sex-working platforms with an "amateurish" flavor, like OnlyFans or Chaturbate. In a web that is already teeming with porn videos, what sex workers have left to sell is a *feeling of intimacy*: they get money to say the (nick)name of a customer aloud, or even to share their phone numbers with them, just for the sake of letting their customers feel closer to them. This craving for intimacy is also likely to underlie the spread of teledildonics, i.e. sex toys remotely controlled by someone else (Liberati, 2017).

NCII can be thought as the dark side of amateur porn. NCII consumers' main craving is not the aesthetic content of the image or video in itself—they could have had access to millions of such contents without committing illegal and immoral actions. Rather, it is the sense of proximity and intimacy. Yet, unlike customers of camshows, who *buy* this feeling of intimacy, people who share and consume NCII illicitly *steal* intimacy. Arguably, this is not in order to save money. Most likely, a major driver of their behavior is the abusive nature of voyeurism and the sense of power resulting from seeing something they were not supposed to see, and that cannot return their gaze[10]. In this paper, however, we cannot and will not diagnose the multi-faceted etiology of IBSA perpetration. Instead, we want to pursue the following hypothesis: insofar as these contents lose their power to express intimacy, NCII will lose part of their allure (and their power to harm), likely resulting in less circulation simply because the abusers will realize that there is no intimacy left to steal.

Recall the two examples cited at the beginning of this paper, namely the hacker's sextortion attempt, and the leakage of Emi Cilibiu's private sex tape resulting in her public shaming and possibly in her getting fired from her job. Would they have had the same bite if they had been performed via other kinds of images than photographs or videos? Compare the threat of seeing a video depicting you performing a sexual act with the same threat made in relation to a comic or a painting representing the same content. While still being rather unpleasant and worthy of legal consideration (see Sect. 6), the latter kind of blackmail intuitively sounds weaker than that based on actual photographs or video footage. And now, imagine a remake of *Bad Luck Banging or Loony Porn* in which, instead of having their video leaked, the protagonists are seen performing sexual activities by a malicious painter, who later draws a disturbingly detailed reproduction of their intercourse. No doubt it might be embarrassing for the protagonists, but could it credibly lead to the same level of shame experienced by Emi after the leakage of her video? Could it be used as leverage by bigoted parents to get her fired? Intuitively, the answer is no. And we now know why, namely because visual representations like hand-made drawings lack the special epistemic and affective power typically possessed by photographs and videos.

But what if photos and videos, no matter how realistic they look, lose this special power, and begin to be perceived as just any other hand-made image?

[10] "The voyeur's pleasure depends on the object of this look being unable to see him: to this extent, it is a pleasure of power, and the look a controlling one" (Kuhn, 1985, p. 28. See also Calvert and Brown, 2000; Henry et al. 2020, ch. 5).

5 Will deepfakes subvert the status (quo) of photographic experience?

To further explore our intuition, let us examine how Walton, when making the case for his thesis that photographs are transparent, imagines what visual experience we would have in front of one of Chuck Close's hyper-realistic self-portraits. At first sight, we may have the impression of staring at a photograph of Close's face (or seeing *Close himself*, if we accept Walton's thesis[11]). But then, at a closer inspection, we realize that the artist has deceived us: what we mistook for a photograph turned out to be a hyper-realistic painting. What happens next? According to Walton (1984, p. 4),

> The discovery jolts us. Our experience of the picture and our attitude toward it undergo a profound transformation, one which is much deeper and more significant than the change which occurs when we discover that what we first took to be an etching, for example, is actually a pen-and-ink drawing. [...] We feel somehow less "in contact with" Close when we learn that the portrayal of him is not photographic.

But let us push this mental experiment a little further. Imagine that this jolt motivates us to dig deeper into Close's artistic production and biography. We learn that mimicking photography with painting is a *leitmotif* of his art. One day, a friend who knows how much we like Close's art invites us to an exhibition of his self-portraits, many of which have never been displayed before. While looking at these masterpieces, we renew our admiration for Close's capacity to recreate in painting the visual effects of photography. One painting strikes us as impressively accurate. "What a skillful artist!", we murmur. But then, after reading the description of the canvas, we realize that what we are staring at is not a painting, after all, but an actual photograph. Again, the discovery jolts us. We had become so used to taking Close's realistic depictions as paintings, despite their striking resemblance to actual photographs, that we failed to detect a real photograph as such. After becoming acquainted with his art, our attitude toward Close's photographic-looking images had ceased to be that of the trustful non-reductionist *à la* Reid and switched to that of the skeptical reductionist *à la* Hume: the fact that the image looked like a photograph did not suffice anymore for us to treat it as such; an additional ingredient was needed (in this case, the description) in order to receive photo-looking images as actual photos.

Our follow-up to Walton's experiment is aimed at suggesting that our doxastic and affective stance toward photographic images may not be carved in stone. Rather, it is at least conceivable that it can be modified given the appropriate circumstances. Many scholars who have elucidated the properties of photography have pointed out that its status is contingent on our psychology or our beliefs about how photos are generated (see notably Cohen & Meskin 2004, Sect. 7). Indeed, according to some, this status is already being disputed. For instance, Savedoff forecasted that "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

[11] But see (Walton, 1984, footnote 29).

could all but disappear" (2000, p. 202; but see Hopkins 2012). Cavedon-Taylor (2013, p. 88) forecasted that, with the development of digital manipulation technologies "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".

Now, if digital photography has ignited a shift in our attitude toward images that *look like* photos, deepfake technology is going to accelerate this shift by allowing cheap mass production of fabricated images that are visually indistinguishable from actual photos. While for photorealistic paintings such as Close's self-portraits mimicking the aesthetic properties of photographs requires a significant amount of skill and effort, producing a deepfake with apps like DeepNude needs but a click and some seconds. Hence, just like getting to know Chuck Close's artistic intents and skill can turn us into skeptics toward realistic images depicting him, we surmise that the increased awareness of the possibilities of deepfake technology could promote skepticism as a default attitude toward (m)any images that look like photographs and videos.

We are not alone in believing this: other philosophers have tackled the epistemic consequences of the spread of deepfakes, focusing on the "threat" they pose to knowledge. Fallis (2021) points out that, by raising the likelihood that what is represented in a video does not correspond to anything that actually happened, deepfake technology will sow distrust toward *all* videos, including those that have been genuinely filmed (after all, how could you be sure that their etiology is reliable?). By so doing, deepfakes will reduce the amount of knowledge we may acquire from videos in general. Rini (2020) highlights that, as deepfakes erode the reliability of videos, they can no longer be used to double-check testimony; nor can they dissuade people from giving false testimony. Hence, in the long run they can no longer be used to level the epistemic playing field of testimony, which is rigged by several epistemic injustices (Fricker, 2007) due to power imbalance[12].

We do not deny that deepfakes may have these and other nefarious epistemic consequences (but see Harris, 2021). Nonetheless, we claim that, once we turn our gaze to NCII and other IBSA, these very epistemic drawbacks could end up yielding some positive outcomes. In a world where most intimate images were *known to be* deepfakes, we would be less worried about what images and videos (including real photographic images and videos) could reveal about us, because hardly anyone would assume by default that they were revealing something about us.

Note, however, that the diffusion of deepfake technology is not sufficient *per se* to warrant the transition to the skeptical attitude we envision. Fallis (2021, fn. 17) correctly notes that "Strictly speaking, what matters is not the probability that the video exists, but the probability that the video is available to be seen". We believe, nevertheless, that a further step is necessary in order to obtain the skeptical shift that he and we are visualizing, namely, that viewers become *aware* of the potential of this technology to fabricate realistic images and videos. Acknowledging this step allows us to appreciate the relevant role that digital literacy may have in promoting the mitigation we are picturing here, but also to alleviate the problem highlighted by Fallis (although not that highlighted by Rini, unfortunately). We will come back to this matter in the final section of this paper (Sect. 6).

[12] While Fallis (2021) and Rini (2020) focus mainly on video, their arguments also extend to still images.

Of course, the issue with IBSA is not just a matter of *doxastic attitudes*. The sense of intimacy elicited by the visual experience of (what we take to be) a photograph or video also plays a major role. Right after presenting the mental experiment regarding Close's self-portrait, Walton noted that "If a painting is of a nude and if we find nudity embarrassing, our embarrassment may be relieved somewhat by realizing that the nudity was captured in paint rather than on film" (1984, p. 4). Again, let us push the idea a little further: if realizing that what we have mistaken for a nude photographic portrait (or video) is actually a hand-made picture (or a fabricated video) can mitigate our feeling of embarrassment, then the same discovery could also reduce sexual arousal, or whatever feeling motivates the perpetration of NCII. And the same mitigation would apply to images or videos that were *actually* photographed or filmed when independent proofs of their etiology are lacking. In a world in which photographs or videos are no longer seen as reliable seals of indexicality because it is safer to assume that they are fabricated rather than captured by a camera, people willing to consensually share and vindicate their naked pictures (e.g. in sexting or exhibitionism) will end up being faced with the issue of certifying their contents. Perhaps they will adopt a blockchain (cf. Floridi, 2018). Indeed, some have already begun to respond to the challenge of providing their photographs with new seals of indexicality. Within Reddit's digital platform, the subreddit Gonewild, designed for the consensual sharing of naked pictures of women, has developed what van der Nagel (2020) calls an "embodied verification system". While posting their photos, users are invited to "crumple the sign up into a ball and then take [their] pictures with the sign uncrumpled [because it] creates a lot of random angles in the paper, and convinces the [moderators] and users that the sign was not photoshopped" (quoted in van der Nagel, 2020).

Again, we are not the first to forecast this depowering of pictures. For instance, focusing on the impact of digital technology upon our cognitive phenomenology, Cavedon-Taylor (2015, p. 88) envisioned that "We may well reach a point at which what we see in photographs we no longer feel in (quasi-)perceptual contact with, at least not in the way that previous generations did".

However, he is also open to the possibility that "insofar as we continue to undergo pictorial experiences before such photographs, some degree of quasi-perceptual phenomenology should be thought to occur" (ibid.). Indeed, we acknowledge that our attitude toward allegedly photographic pictures or videos may be partly driven by some hardwired psychological disposition, rather than by our beliefs about the genesis of photographic images/videos. For instance, Ferretti (2018) suggested that the key ingredient for the *visual feeling of presence* is a property afforded by binocular vision, i.e. qualitatively rich stereopsis, which allows us to estimate egocentric depth based on the comparison between the visual information of the two retinas, namely the kind of information that artists skillfully manipulate when they paint *trompe l'oeils*. Similarly, Walden (2016) makes a strong case that the feeling of affective contact is due to a deep similarity between the workings of our visual system when we attend to pictures and during unmediated vision. He argues that, contrary to our epistemic attitude, which may be overturned by background knowledge, the affective sense of contact is unaffected by beliefs. In fact, Walden (2016, p. 48) reports that his own phenomenology "cannot concur with [Walton's] report that there is a lessening of perceptual contact with the youthful Close upon making the discovery" that his self-portrait is not a photo.

Anscomb (2022) concurs with Walden that our affective reaction to picture (unlike the epistemic one) is driven by the workings of our visual system and is quite immune to the influence of beliefs. Nevertheless, she notices that beliefs can exert an indirect effect on our affective reaction by mobilizing our visual attention, for instance in order to reduce the cognitive dissonance between a feeling of affective contact toward a certain image and the belief that it has not been produced indexically. For instance, were we approaching Close's self-portrait knowing it is not a photograph, even if we would still feel intimate with him at first (as predicted by Walden), she suggests that our knowledge is likely prompt us to visually inspect the painting until we find some detail that gives away its non-indexical nature, hence lessening our feeling of contact (see Anscomb 2022, Fig. 2).

Qualitative evidence collected by Flynn et al. (2021) provide some ground to suspect that some cognitive dissonance may occur between the feeling of affective proximity elicited by a fake NCII and the belief that it is forged[13]. Yet, whether our affective reaction toward images and videos is indeed driven by low-level cognitive machinery rather than on culturally acquired habits is still an open question, which must be addressed empirically. Based on our current knowledge, we cannot exclude that the first impact with intimate images may preserve some of its affective allure even in a world where deepfakes are the norm—just like *trompe l'oeils* can preserve some of their illusionary depth even after we discover the trick and staring at Chuck Close's self-portraits may make us feel in contact with him even if we know they are not photographs.

Would the belief-insensitivity of our feeling of intimacy undermine our qualified optimism? We do not think so. Certainly, it is still possible that the thrill felt during those initial seconds could motivate the consumers of deepfake NCII to persist in producing and sharing them. And yet, within a sufficiently skeptical social environment, experiencing intimacy with a NCII that is very likely forged will place a significantly higher burden on their suspension of disbelief. Recall that IBSA perpetrators are not after the aesthetics properties of pictures (they could easily find plenty of them in legal venues). Instead, they seek to steal intimacy with *actual* victims. And indeed, they sometimes gather other information about them in order to strengthen and restore the credibility and feeling of contact with them (a practice called 'doxing'; see Sect. 6).

To sum up, in the present section we have argued that the special status of our "trustful and intimate" visual experience of (what we take to be) photographs or videos, which we described in the previous section, can be subverted toward a more "skeptical and detached" attitude; and that people's heightened awareness of the potentialities of deepfake technologies will facilitate this transition from the former to the latter attitude. Hence, the "epistemic maelstrom" (Rini, 2020, p. 8) afforded by deepfakes

[13] Notably, upon finding that a perpetrator has generated NCII of her via deepfake, one victim shifted from identifying to de-identify herself with the depicted woman: "[He] had an entire folder of photos on his desktop of people who looked remotely like me. … The first ones I saw … I thought they were me, and then I had a closer look, and I was like, hang on a minute, I have marks on my body that aren't there. … So yeah, my first reaction was: 'Oh my god, this is a thing that happens to other people that's suddenly happening to me'. And then my second was, 'oh that's a relief, it's not actually me'. And then it was like, 'oh no, that's me' ". (Flynn et al., 2021, p. 6).

may end up having a positive consequence, i.e. mitigating the consequences of a societal maelstrom such as that of IBSA.

Does it follow that we can stop worrying about NCII and IBSA? Unfortunately, things are more complicated than this, and several caveats must be considered before drawing any implications from our arguments. We discuss these in the next and conclusive section.

6 Conclusive remarks, caveats, and some implications

Our discussion began with an attempted scam based on the threat to share some intimate video footage that a hacker alleged to possess. It proceeded by mentioning the misfortunes of Emi Cilibiu, the protagonist of *Bad Luck Banging or Loony Porn*, whose private video went viral, resulting in public humiliation and possibly in the loss of her job. These are but two examples of widespread and problematic phenomena: the diffusion of Non-Consensual Intimate Images (NCII) and other Image-Based Sexual Abuses (IBSA). We provided some context to understand the social unfolding of these phenomena (Sect. 2), and then (in Sect. 3) asked how these social problems will be impacted by the rapid spread and sophistication of technologies that allow for the cheap production of reliable deepfakes. Contrary to the most pessimistic predictions, we offered an argument for qualified optimism, hinging on two premises. First: an important drive motivating perpetrators of IBSA is the allure that photographs and videos have in virtue of their special epistemic and affective status, which makes us feel especially trustful and intimate toward them (Sect. 4). Second: the awareness of the potential of deepfakes is going to disrupt this allure, shifting our default attitude toward (what we take to be) photographs and videos, possibly including genuine ones, toward a skeptical and (possibly) detached attitude, thus diminishing their allure (Sect. 5).

While we are confident that our argument offers grounds for *some* optimism, it does not warrant naive and unconstrained optimism. All the less does it warrant a laissez-faire attitude toward NCII and other IBSA.

A reader who accepts our argument may be tempted to undertake the following reasoning: "if spreading deepfakes is likely to accelerate the loss of these epistemic and affective properties that make them a dangerous tool for IBSA, why not accelerate the process by actively spreading deepfakes?" A similar approach is being taken by scholars who, following the metaphor of vaccines, employed the *inoculation* of 'inactivated' fake news to raise subjects' defenses against online manipulations (see Lewandowsky & van der Linden, 2021).

However, it should be kept in mind that, as highlighted by several scholars, deepfake intimate images can still harm *even if manifestly false*. In fact, even if their nature of fabricated images is clear to the onlooker, in the light of the highly gendered-biased context in which most of them arise (Semenzin & Bainotti, 2020; Henry et al., 2020) they nonetheless reproduce despicable practices of exploitation and sexualization of the female body (Öhman, 2020) and promote sexual objectification via implicit psychological association (Harris 2021), as does the sexualizing male gaze in films (Mulvey, 1975). While many of these concerns also apply to other forms of

hand-made images (e.g. erotic comics featuring an existing, non-consensual victim), deepfake-generated NCII may require special attention by virtue of the psychological implications of the profound realism that makes them potentially "more gripping" (Rini, 2020, p. 11; see also De Ruiter, 2021), especially if turns out that knowing that a picture is not a genuine photograph does not automatically dampen the feeling of intimacy toward it (as claimed by Walden, 2016).

Moreover, notice that the skeptical scenario we are envisioning will only unfold in the long run. Just as our current "Reidian" attitude toward photographic images is the fruit of decades of familiarization with the nuances of photographic technology, so it is likely that switching to a skeptical attitude may be a matter of decades. But echoing a famous Keynesian adage, "The *long run* is a misleading guide to current affairs. In the *long run* we are all dead". In other words, in the wake of the skeptical turn that will mitigate the allure of NCII, we should not reduce our efforts to protect the victims and pursue the perpetrators. We cannot exclude that, although the spread of deepfakes will ultimately end up depowering them, during the intermediate steps they may cause a lot of harm, especially if the availability of deepfake technologies spreads faster than the awareness of their workings.

This brings us to our third remark. It should be remembered that the diffusion of deepfakes *per se* will not suffice to disrupt the special status of photographs and videos: the awareness of their workings is also necessary. However, while we are confident that this awareness will increase overall in society, it is highly unlikely that this will happen evenly for everybody. More likely, it will affect some individuals more than or before others, depending on several factors, such as digital literacy or age. Thus, like other scholars before us (e.g. Wagner & Blewer, 2019; Naezer & van Oosterhout, 2021), we call for digital education programs, and especially sensitization toward the possibility of image and video manipulation. An intriguing possibility would be to attempt controlled and unharmful expositions to fake contents, similarly to what happens with vaccines (cf. Lewandowsky & van der Linden, 2021). In this respect, artists and art curators whose work aims at exposing the non-indexical nature of photographs goes in the right direction (for a notable example, see Fineman, 2012). Such programs can also shield society from other epistemic damage brought about by deepfakes, e.g. in politics and other societal issues. Yet, it is not prudent to expect digital literacy to be a *panacea*[14]. In fact, it is entirely possible that our attitude toward alleged photographs or videos is not governed so much by our current conscious beliefs about their productions, but rather by *habits* that we have learnt through life, and which we can hardly *un*learn. If that is the case, we should be prepared for the possibility that older generations may never reach the same level of disillusion as digital natives. And

[14] May digital literacy be even detrimental in some cases? Perhaps, when someone learns about the etiology of deepfakes, they may regain some of the special epistemic and affective status of photographs, as they are based on a database of *actual* pictures. We think that this objection is not so troubling for our account. Indeed, the referential status of NCII deepfakes resembles that of a composite picture made up by photographs of body parts of several persons (although their computational nature pushes our puzzle further, as it does not mix body parts but rather abstract variables). The resulting depicted person would thus be a fictional character. And while someone may wish to feel intimate with fictional characters (e.g. those who seek for erotic comics or overt deepfakes), the motives of IBSA perpetrators (e.g. revenge, control, humiliation; see Henry et al., 2020, ch. 4) often requires that the NCII refers to specific and actual persons.

in any case, we should expect friction deriving from different people having different epistemic and affective attitudes in the face of images.

Indeed, we have reason to think that many people are already on the route toward the skeptical and detached attitude we envision. As younger generations routinely share selfies of themselves with filters that modify the way they look, often in realistic ways, they are likely to have been inoculated into disbelieving that what they see corresponds to reality. Indeed, based on 12 focus groups involving young *Canadians* (18–30 years), Lavrence & Cambre (2020) report that participants adopt a 'digital-forensic gaze' toward selfies, which recalls Anscomb's (2022) description of the skeptic onlooker faced with Close's self-portratis: "when looking at selfies, the use of filters was usually presumed, and deciphering authenticity is integral to what drives looking practices" (p. 5)[15]. As Harris puts it, although "the photograph is weaker evidence than it would have been prior to the popularity of Photoshop" (2021, p. 7), we have not witnessed anything like the epistemic catastrophe envisaged by some epistemologists, because people have developed epistemic antibodies against deception, e.g. checking the source or using blockchains. In a not-so-distant future, people engaging in sexting may be required to provide independent seals of indexicality like those employed in the subreddit Gonewild! (van der Nagel, 2020).

Alas, such extra-photographic seals will not only be available to consensual adults: the perpetrators of IBSA may also find other means to bypass the blockade of skepticism, thus restoring the credibility and the affective allure of NCII. Extant analyses of websites and private chatrooms where NCII are spread show that they are often accompanied by personal information about the victims (*doxing*), links to their social media, and private anecdotes motivating why the NCII has been posted in cases of classical 'revenge' scenarios (Hearn & Hall, 2019; Semenzin & Bainotti, 2020; Uhl et al., 2018). We predict that, as images and videos per se will see their credentials eroded by the increased awareness of the potential of deepfakes, IBSA perpetrators will increasingly rely on such supplementary information. This could result in new risks for the victims, but also new possibilities for persecuting the perpetrators on the grounds of privacy violations.

While the list of caveats reported above may be non-exhaustive, they alone should suffice to stress how the implication of our argument is *not* that we, as society, can lower our guard with respect to NCII and IBSA because the problem will auto-dissolve. Rather, we must seek to better understand the phenomenon so as to predict its trajectory and fight it properly. If we are right, distrust and detachment toward images will prove a powerful ally in this endeavor.

Acknowledgements This paper is the fruit of the joint effort and discussion of both authors. In particular, MV drafted the first manuscript, CV proposed substantial revisions, and both approved the final version. The authors are grateful to Marco Facchin, Massimo Leone, Neri Marsili, Maria Oliva, Irene Papa, and Alberto Romele for their valuable feedback on the manuscript. The feedback provided by two anonymous referees was tremendously helpful to ameliorate the paper. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No 819649-FACETS).

[15] Lavrence and Cambre also suggest a possible tension between the epistemic and affective response to photographs: "although the digital-forensic gaze is skeptical about what it sees, on another level, it takes what it sees at face-value and feels the image as if it were real" (Lavrence and Cambre, 2020, p. 9).

Funding Open access funding provided by Università degli Studi Roma Tre within the CRUI-CARE Agreement. This study was supportede by HORIZON EUROPE European Research Council project FACETS (Grant No. 819649).

Declarations

Conflict of interest The authors have no conflict of interest to declare.

References

Anscomb, C. (2022). Look a little (Chuck) closer: aesthetic attention and the contact phenomenon. *British Journal of Aesthetics*, *62*(3), 475–492.

Byron, P., McKee, A., Watson, A., Litsou, K., & Ingham, R. (2021). Reading for realness: Porn literacies, digital media, and young people. *Sexuality & Culture*, *25*(3), 786–805.

Caletti, G. M. (2021). Can affirmative consent Save" revenge Porn" laws? Lessons from the italian criminalization of non-consensual pornography. *Va JL & Tech*, *25*, 112.

Calvert, C., & Brown, J. (2000). Video Voyeurism, privacy, and the internet: Exposing peeping toms in cyberspce. *Cardozo Arts & Ent. LJ, 18*, 469.

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

Cavedon-Taylor, D. (2015). Photographic phenomenology as cognitive phenomenology. *British Journal of Aesthetics*, *55*(1), 71–89.

Chesney, B., & Citron, D. (2019). Deep fakes: a looming challenge for privacy, democracy, and national security. *California Law Review*, *107*, 1753.

Cohen, J., & Meskin, A. (2004). On the epistemic value of photographs. *The Journal of Aesthetics and Art Criticism*, *62*(2), 197–210.

Cole, S. (2017). AI-assisted fake porn is here and we're all fucked. *Motherboard (tech by Vice)* December 12 https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn. Accessed 15 July 2022.

Costello, D., & Phillips, D. M. (2009). Automatism, causality and realism: foundational problems in the philosophy of photography. *Philosophy Compass*, *4*(1), 1–21. https://doi.org/10.1111/j.1747-9991.2008.00193.x.

Currie, G. (1999). Visible traces: Documentary and the contents of photographs. *The Journal of Aesthetics and Art Criticism*, *57*(3), 285–297.

De Ruiter, A. (2021). The distinct wrong of deepfakes. *Philosophy & Technology, 34*(4), 1311–1332.

Fallis, D. (2021). The epistemic threat of deepfakes. *Philosophy & Technology*, *34*, 623–643.

Ferretti, G. (2018). Visual feeling of presence. *Pacific Philosophical Quarterly*, *99*, 112–136.

Fineman, M. (2012). *Faking it: manipulated photography before photoshop*. New York: Metropolitan Museum of Art.

Floridi, L. (2018). Artificial intelligence, deepfakes and a future of ectypes. *Philosophy and Technology, 31*(3), 317–321.

Flynn, A., Powell, A., Scott, A. J., & Cama, E. (2021). Deepfakes and digitally altered imagery abuse: a cross-country exploration of an emerging form of image-based sexual abuse. *The British Journal of Criminology*. https://doi.org/10.1093/bjc/azab111.

Fricker, M. (2007). *Epistemic injustice: power ad the Ethics of Knowledge*. Oxford University Press.

Friday, J. (2002). *Aesthetics and photography*. Ahsgate.
Geen, R. G. (1975). The meaning of observed violence: real vs. fictional violence and consequent effects on aggression and emotional arousal. *Journal of Research in Personality*, *9*(4), 270–281.
Goodfellow, J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, J. (2014). Generative adversarial nets. *Advances in neural information processing systems*. https://doi.org/10.48550/arXiv.1406.2661
Gosse, C., & Burkell, J. (2020). Politics and porn: how news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication*, *37*(5), 497–511.
Harrington, C. (2021). What is "Toxic Masculinity" and why does it matter? *Men and Masculinities, 24*(2), 345–352.
Harris, K. R. (2021). Video on demand: what deepfakes do and how they harm. *Synthese*, *199*(5), 13373–13391.
Hearn, J., & Hall, M. (2019). 'This is my cheating ex': gender and sexuality in revenge porn. *Sexualities*, *22*(5–6), 860–882.
Henry, N., McGlynn, C., Flynn, A., Johnson, K., Powell, A., & Scott, A. J. (2020). *Image-based sexual abuse: a study on the causes and consequences of non-consensual nude or sexual imagery*. Routledge.
Hopkins, R. (2012). Factive pictorial experience: What's special about photographs? *Noûs, 46*(4), 709–731.
Jochelson, R., Ireland, D., & Taylor, H. (2021). Clearing your history: a review of non-consensual distribution of intimate images in Canada and future responses. *UBCL Rev*, *54*, 763.
Kuhn, A. (1985). *The power of the image. Essays on representation and sexuality*. Routledge.
Lavrence, C., & Cambre, C. (2020). Do I look like my selfie?": Filters and the digital-forensic gaze. *Social Media + Society*. https://doi.org/10.1177/2056305120955182
Leonard, N. (2021). Epistemological problemsof testimony. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy.*
Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, *32*(2), 348–384.
Liberati, N. (2017). Teledildonics and new ways of "being in touch": a phenomenological analysis of the use of haptic devices for intimate relations. *Science and engineering ethics*, *23*(3), 801–823.
Maddocks, S. (2018). From non-consensual pornography to image-based sexual abuse: charting the course of a problem with many names. *Australian Feminist Studies*, *33*(97), 345–361.
Maes, C., & Vandenbosch, L. (2022). Physically distant, virtually close: adolescents' sexting behaviors during a strict lockdown period of the COVID-19 pandemic. *Computers in Human Behavior*, *126*, 107033.
McGlynn, C., Rackley, E., & Houghton, R. (2017). Beyond 'revenge porn': the continuum of image-based sexual abuse. *Feminist Legal Studies*, *25*(1), 25–46.
Meskin, A., & Cohen, J. (2010). Photographs as evidence. In S. Walden (Ed.), *Photography and philosophy: Essays on the pencil of nature* (pp. 70–90). Wiley.
Mulvey, L. (1975). Visual pleasure and Narrative Cinema. *Screen*, *16*, 30–52.
Naezer, M., & van Oosterhout, L. (2021). Only sluts love sexting: Youth, sexual norms and non-consensual sharing of digital sexual images. *Journal of Gender Studies*, *30*(1), 79–90.
Öhman, C. (2020). Introducing the pervert's dilemma: A contribution to the critique of Deepfake Pornography. *Ethics and Information Technology, 22*(2), 133–140.
Paasonen, S. (2010). Labors of love: netporn, web 2.0 and the meanings of amateurism. *New Media & Society*, *12*(8), 1297–1312.
Peirce, C. S. (1984). What is a sign? Reprinted in: Peirce, C. S. (1982). The writings of Charles S. Peirce: A chronological edition. Volumes 2. *Peirce edition project.* Indiana University Press.
Pettersson, M. (2011). Depictive traces: on the phenomenology of photography. *The Journal of Aesthetics and art criticism*, *69*(2), 185–196.
Powell, A., & Henry, N. (2014). Blurred lines? Responding to 'sexting'and gender-based violence among young people. *Children Australia*, *39*(2), 119–124.
Rini, R. (2020). Deepfakes and the epistemic backstop. *Philosophers, 20*(24), 1–16.
Sadowski, P. (2011). The iconic indexicality of photography. In P. Michelucci, C. Ljungberg, & O. Fischer (Eds.), *Semblance and signification* (pp. 355–368). John Benjamins.
Savedoff, B. E. (2000). *Transforming images: how photography complicates the picture*. Cornell University Press.

Semenzin, S., & Bainotti, L. (2020). The use of telegram for non-consensual dissemination of intimate images: gendered affordances and the construction of masculinities. *Social Media + Society*, *6*(4), 2056305120984453.

Thomas, M. H., & Tell, P. M. (1974). Effects of viewing real versus fantasy violence upon interpersonal aggression. *Journal of Research in Personality*, *8*(2), 153–160.

Uhl, C. A., Rhyner, K. J., Terrance, C. A., & Lugo, N. R. (2018). An examination of nonconsensual pornography websites. *Feminism & Psychology*, *28*(1), 50–68.

van der Nagel, E. (2020). Verifying images: Deepfakes, control, and consent. *Porn Studies, 7*(4), 424–429. https://doi.org/10.1080/23268743.2020.1741434

Wagner, T. L., & Blewer, A. (2019). "The word real is no longer real": Deepfakes, gender, and the challenges of ai-altered video. *Open Information Science, 3*(1), 32–46.

Walden, S. (2016). Transparency and two-factor photographic appreciation. *British Journal of Aesthetics*, *56*(1), 33–51.

Walton, K. L. (1984). Transparent pictures: On the nature of photographic realism. *Critical Inquiry, 11*(2), 246–277.