



Introducing the pervert's dilemma: a contribution to the critique of Deepfake Pornography

Carl Öhman¹

Published online: 19 November 2019
© The Author(s) 2019

Abstract

Recent technological innovation has made video doctoring increasingly accessible. This has given rise to Deepfake Pornography, an emerging phenomenon in which Deep Learning algorithms are used to superimpose a person's face onto a pornographic video. Although to most people, Deepfake Pornography is intuitively unethical, it seems difficult to justify this intuition without simultaneously condemning other actions that we do not ordinarily find morally objectionable, such as sexual fantasies. In the present article, I refer to this contradiction as the *pervert's dilemma*. I propose that the method of Levels of Abstraction, a philosophical mode of enquiry inspired by Formal Methods in computer science, can be employed to formulate at least one possible solution to the dilemma. From this perspective, the permissibility of *some* actions appears to depend on the degree to which they are abstracted from their natural context. I conclude that the dilemma can only be solved when considered at low levels of abstractions, when Deepfakes are situated in the macro-context of gender inequality.

Keywords Deepfake · Pornography · Pervert's dilemma · Gamer's dilemma · Information ethics

Introduction

Artificial intelligence (AI), and in particular so-called Deep Learning algorithms, provide users the flexibility to edit and manipulate digital video content. Similar technologies are widely used on popular apps like Snapchat (and Instagram), which has a Face Swap feature that allows users to switch faces with one another in live videos. They have also provided Hollywood filmmakers the ability to add deceased actors, such as Peter Cushing and Oliver Reed, into new movies (Minton 2017). But Deep Learning is now increasingly used for another purpose: to generate pornographic content commonly known as *Deepfake Pornography*.

Deepfakes broadly refer to hyper-realistic videos in which a person's face has been analysed by a Deep Learning algorithm, and then superimposed on top of the face of an actor in a video. Since the algorithm has "learned" the face's features from different angles, and how it moves in different expressions, it can replicate it in a way that follows the expressions of the actor. To clarify, this does not necessitate

any privacy infringement or illicit information access (Harris 2019). It can be done with publicly available pictures or video material. Much like a human brain, the Deep Learning algorithm "learns" from the informational input it is fed and is then able to generate its own amalgamation of it. In this respect, there is, on a conceptual level, little difference between a picture created by a Deep Learning algorithm and a picture one can imagine in one's head based on what one has seen. Thus it is not unlike an artificial, or at least an augmented, fantasy.

The Deepfake phenomenon first emerged in 2017 and exploded in sophistication and popularity during early 2018 (Cole 2018). The launch of programs like FaceApp made it possible for amateurs and enthusiasts to create their own Deepfake videos using the app and a piece of video material of the person whose face they were interested in using. As one may expect, the technology—previously only accessible to Hollywood CGI experts—is now mainly used to create pornographic videos starring female celebrities such as Gal Gadot and Emma Watson. But since the software works just as well with input data from platforms such as Instagram and YouTube, it is reportedly also used to create content based on the faces of ex-girlfriends and mere acquaintances (Harwell 2018). Whereas this theoretically makes anyone a potential target of Deepfake Pornography, the phenomenon

✉ Carl Öhman
carl.ohman@oii.ox.ac.uk

¹ Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK

so far appears to be heavily gendered. Like most pornographic content, it is predominantly produced by and for a male audience, although this time (fictionally) starring women who have not given their consent. Sites such as Pornhub, Reddit, and Twitter have thus banned deepfake content, which has led to the launch of number of new sites that are specially devoted to sharing, creating, and teaching users how to make their own Deepfakes.

There is certainly much to be said about Deepfakes, both from a political, legal and ethical point of view. In this essay, however, I shall focus only on a specific moral dilemma that arises from the phenomenon, which I shall refer to as the *pervert's dilemma*, for lack of a better term.¹ Although Deepfake pornography (henceforth just “Deepfakes”) strikes most people as intuitively disturbing and immoral—recall that several sites (e.g., Reddit, Pornhub, etc.) pre-emptively banned Deepfakes—it seems difficult to justify this intuition without simultaneously disapproving of other actions not normally considered harmful. For instance, we may again compare Deepfakes to sexual fantasies. Both fantasies and Deepfakes are arguably no more than a virtual image generated by informational input that is publicly available, and thus it is hard to identify a quality that makes the former more permissible than the latter. Yet, although certain sexual fantasies can be deemed impermissible due to the grotesque or violent nature of their content (more on this below), they are not normally considered unethical per se.

It is tempting to argue that creating a Deepfake requires more labour and thus more ill intent than the fantasy. But if this were true, then any sexual fantasy requiring significant labour would be as impermissible as the Deepfake, and this does not sound right. Moreover, I believe our intuitions about Deepfakes would remain even if they could be generated by a simple click of a button (which is virtually the case already). Neither can Deepfakes be condemned with reference to their materiality (the materiality of a fantasy can be debated), as it is hard to see how materiality in and of itself carries any ethical significance. One could of course argue that material objects are more shareable and that this implies at least the potential to ruin the public image of the person it depicts. But this objection, I believe, does not fully capture

our moral intuitions. Even if a Deepfake was not sharable, we would still question its moral permissibility. Consider, for instance, the following example.

A uploads a self-depicting video on some form of social or public media. B then uses these pictures as inputs to a Deep Learning algorithm, despite knowing that A would disapprove of such action. The algorithm analyses the movement patterns of A's face in such a way that it can create a realistic superimposition of it onto that of an actor in a pornographic video. Let us then further add two conditions: The technology used by B guarantees that (i) A can never find out about the pornographic content in which A's face is starring; and (ii) it is impossible to distribute the content to anyone else. These two conditions should prevent any arguments based on A's personal wellbeing or reputation, thus making the materiality of the content morally irrelevant (at least insofar as I can see). Still, I claim, the moral intuition of most people is that B is doing something wrong, despite there not being any immediately identifiable and morally relevant difference between this case and a mere vivid sexual fantasy. Herein lies the dilemma,² which can now be fully articulated thus:

1. Creating pornographic Deepfake videos based on someone's face (without their explicit consent) is morally impermissible.
2. Having private sexual fantasies about someone (without their explicit consent) is per se normally morally permissible.
3. Under conditions (i) and (ii), there is no morally relevant difference between creating a Deepfake video based on someone's face and having a private sexual fantasy about someone.

To prevent misunderstandings, 2 must be further clarified. Sexual fantasies are a rather broad concept involving a number of different subcategories. For instance, Smuts (2016) distinguishes between mere fantasizing, engaging with fictions, and dreaming, arguing that each activity has different moral characteristics, such as the degree to which one pictures oneself as involved in an action, and the degree to which it is voluntary. While some philosophers hold all such activities to be immune to moral criticism (Cooke 2014), others, such as Bartel & Cremaldi (2018), instead argue that fantasies can be morally objectionable insofar as they cultivate desires or pro-attitudes that themselves are morally objectionable—such as a desire to rape (Kershnar 2005).

¹ Note that the connotations to the word “pervert” differ between UK and American English. According to the Cambridge Dictionary (<https://dictionary.cambridge.org/dictionary/english/pervert>), the UK definition of a pervert is “a person whose sexual behaviour is considered strange and unpleasant by most people” whereas the American version reads “a person whose sexual behaviour is considered unnatural and morally wrong”. The difference lies in that the UK definition does not deem the pervert amoral, but merely socially “strange” and “unpleasant” from the standpoint of the majority, while the American definition adds a moral dimension to his or her desires. Writing from the UK, I use the former definition, which applies also to people not doing anything morally (although perhaps socially) wrong.

² It should be noted, however, that this claim is not backed up by any data. Some may find B's actions completely permissible, in which case there is nothing to quarrel about.

We will have reason to revisit these arguments towards the end of this essay, but for now, let us merely state that, even if certain types of fantasies can be considered impermissible, there appears to be a consensus (at least among secular philosophers) that mainstream, everyday sexual fantasies are permissible (see for instance Neu 2012 and Kershnar 2005 supporting this position). Deepfakes on the other hand, I claim, are intuitively impermissible regardless of the permissibility of the acts they depict. To be clear, the contradiction of the pervert's dilemma is thus not that sexual fantasies never can be impermissible, while Deepfakes are always impermissible, but rather that a representation that would (normally) be deemed permissible as a fantasy is deemed *impermissible* as a Deepfake, despite the absence of any immediately identifiable and morally relevant distinction between the two formats. Thus, given that one accepts 1 and 2, it seems that one must either accept that Deepfake content is morally acceptable as long as conditions (i) and (ii) are fulfilled *or* accept that sexual fantasies are morally objectionable despite not directly harming anyone. Neither option seems intuitively right.

In the following section, I shall propose the method of Levels of Abstraction as a means to approach the dilemma. I will show how this method can be employed to produce at least *one* possible response to the pervert's dilemma, in which the morally relevant distinction between Deepfakes and fantasies depend on the degree to which they have been abstracted from their natural context. This strategy allows us to formulate a response whereby Deepfakes are deemed permissible when considered as isolated cases, but impermissible when considered as a phenomenon, whereas sexual fantasies normally appear equally permissible on both levels. Towards the end of the essay, I shall discuss in brief the outlooks for applying my approach also to other similar ethical dilemmas such as the gamer's dilemma introduced by Luck (2009).

The method of levels of abstraction: formalising "it depends"

The pervert's dilemma is inevitably induced by the emergence of sophisticated information technology. For this reason, it makes sense that our approach to unpacking the problem should also take an informational viewpoint. That is, we should understand *A*, *B*, and their actions, as agents acting in response to some kind of informational environment. From this perspective, the question to ask becomes: what type of information is relevant in making a moral judgement regarding the pervert's dilemma? Or better, a more formalised way of asking this question is: what is the relevant *Level of Abstraction* (LoA) for approaching this problem?

The method of LOA is a philosophical mode of inquiry developed by Floridi (2008) with inspiration from Formal Methods in Computer Science. A LoA refers to the extent to which an entity has been "abstracted" from its natural unique context. A person, with her almost infinite complexity, can for instance be reduced to her physical attributes. At this level, in turn, we may introduce a number of *variables*, such as height *h*. When variable *h* is defined using say, the metric system, it becomes an *observable*, something we can measure and use as a means to compare the height of different persons. A LoA can thus be described as a collection of *observables*, that is a set of "possible values and outcomes" (Floridi 2013, p. 31) that enables comparison between entities, be it technologically, morally (e.g. alternative moral actions) or logically.

This is basically just to say that without a common frame of reference, a specification as to what information is relevant, it is impossible to make a comparison. Since an entity consists of an enormous number of possible data, Alice can be a mother, a waitress, an American and a human, and depending on the LoA, some of these will be relevant and others will not. On the LoA of Family Relations, "mother" becomes a relevant observable; on the LoA of Career it is more relevant that she is a waitress. It follows that higher LoAs allow for broader generalization, since the particularities of the analysed system have been reduced. On lower levels, however, generalization is much more difficult since each case has its unique properties. This means that two entities may be the same or different, depending on the LoA we apply. On the LoA of Species, there is no difference between Alice and Bob. On the LoA of Career (lower than species), on the other hand, they may differ. Consider for instance the following example given by Floridi (2011, p. 553):

Whether a hospital transformed now into a school is still the same building seems a very idle question to ask, if one does not specify in which context and for which purpose the question is formulated, and therefore what the required *observables* are that would constitute the right LoA at which the relevant answer may be correctly provided. If the question is asked in order to get there, for example, then the relevant observable is "location" and the answer is yes, they are the same building. If the question is asked in order to understand what happens inside, then "social function" is the relevant observable and therefore the answer is obviously no, they are very different.

The difference between any two things thus depends on which observables we choose to focus on. Note, however, that the method of LoA is in no way a relativist approach. A question is always asked for a purpose—a request for some specific information—and for that specific purpose, there are more or less appropriate LoAs. For instance, the true answer

to the question “Is this the hospital?” is very different for someone in need of a doctor than for someone interested in nineteenth century architecture. This is because a different LoA is required in order to generate a proper response, i.e. different observables come into question. The same principle applies when it comes to moral judgments. Two options may seem equally permissible on one LoA, but different on another. Let me provide an example:

Consider the question of whether it is morally permissible for Alice to break a strike. At the LoA of Nationality (Alice as a citizen of her country), she should arguably break the strike to get industry rolling again; but at the LoA of Class (Alice as a member of a union), she should not. Then again, at the LoA of Family (Alice as a mother) she is morally obligated to break the strike so that she can feed her children. Wittgenstein famously pointed out that we will not find the “real” artichoke by peeling of its leaves (1958, §164). Likewise, we will not find Alice’s “real” obligation regarding her strike by stripping her of all her roles (mother, worker, citizen), that is, on a very high LoA. It is only in her capacity of such roles that she has any moral obligations in the first place. Some actions, such as murder, can be morally evaluated at a very high LoA. Given that we know that it is indeed a case of murder and not manslaughter or mere self-defence, we need to know very little else in order to state that murder is wrong, because it is wrong almost independently on its context. But other actions, or aspects of actions, require a much lower LoA to qualify for ethical evaluation.

To further illustrate the importance of “roles” (i.e. observables) in moral judgements, let us consider another example: is it morally impermissible for Alice to call Bob the N-word behind his back? I believe most people would require more information before they responded to this question. In this case, the relevant LoA is undoubtedly *race*. If Alice is white and Bob is black, then the answer to the question is *yes*. However, if Alice also happens to be black, then the answer is probably *no*. The moral status of the action in question thus depends on the social relations between the categories (observables) at the LoA in question; not so much on the relationship between Alice and Bob as individuals, but on the relationship between the societal groups to which they belong. In the present case, the history of slavery and racism simply cannot be subtracted when making a moral judgement. Even though it may not harm Bob as an individual, most people would agree that it is bad for black people as a collective identity to be referred to in such terms.

Now consider the case of hate crimes. A hate crime consists of two types of harm, one which is directed to the individual who is immediately harmed by the action, and one which is directed towards the group or collective identity of which the individual is part. While the former is prevalent on very high LoAs, the latter can only be detected at a lower LoA. Moreover, an action fails to produce the former may

in certain instances still lead to the latter. Corvino (2002, p. 218) provides an illuminating real-life example:

Some years ago I attended a large Southern university where one of the local fraternities annually held an “Old South Ball.” The fraternity, which was notorious for its white-only membership, would hire black students to pose as “slaves” at the ball for the sake of verisimilitude. Needless to say, this event regularly provoked a serious outcry within the campus community. While some defended the fraternity on the grounds that the black actors were willfully (though, to many minds inexplicably) participating, most thought that the event involved a serious failure on the part of all participants to adopt an appropriate attitude toward slavery. The fact that these actors were paid well was beside the point.

While the Old Southern Ball failed to produce the first type of harm mentioned above, it surely produced the latter, and anyone who fails to appreciate this also fails to make an adequate moral assessment. What Corvino describes is a clash between two LoAs—one which focuses on the individuals involved and one which focuses on the relationship between the collective identities involved. Both sides are right, but the latter level is arguably more relevant because it engages more ethically adequate observables. The lower LoA here contains what Partridge (2011, p. 307) calls an *incorrigible social meaning*. That is, the “range of reasonable interpretations” is limited so that “anyone who has a proper understanding of and is properly sensitive to the moral landscape” will find it objectionable. In the case of the Old Southern Ball, the proper understanding of the moral landscape is that which considers the harm that arises from a *system* of actions, rather than a series of isolated events.

Essentially, this is saying that the ethical significance of the totality of a series of actions *may* in some cases amount to more than the sum of its individual parts. A more formalised way of expressing the same argument is through the concept of Distributed Morality (DM) (Floridi 2012), which analyses ethics from the viewpoint of Multi-Agent Systems (MAS). A MAS is an assemblage of several human actors, machines, virtual environments and even mere concepts. Because of the distributed nature of the system, it may be difficult to allocate the responsibility when it comes to the consequences of the MAS working as a unit. To describe this, DM draws inspiration from distributed knowledge in epistemology. Floridi (2012, p. 729) provides an illuminating example:

Consider the case in which A knows only that $[P \vee Q]$, e.g. that “the car is in the garage or Jill got it”, whereas B only knows that: P, i.e. that “the car is not in the garage”. Neither A nor B knows that Q, only

the supra-agent (with “supra” as i “supranational”) $C = A \cup B$ knows that Q . It is the aggregation of A 's and B 's epistemic states that leads to C knowing that Q .

The same logic applies to morality. That is, although its components may be individually morally permissible, Q can still be morally *impermissible*. The actions of agent A and B can both be neutral, yet their consequences devastating. For example, (at least under appropriate circumstances of pressure and gravity), fire is the direct sum of fuel, oxygen, and heat combined. Yet the damage caused by a fire is not the sum of the damage of fuel, oxygen and heat in isolation. Thus, when we consider the morality of an action, we must place focus also on the system in which this action takes place—the lower LoA. Lighting a cigarette may be disastrous if you are at a gas station, yet the isolated action is per se (relatively) harmless. In some cases, it may be impossible to isolate the role of a single unit in building the totality, (a so called Sorites paradox). For instance, 100,000 grains of sand is certainly a heap, and removing one grain does not change that. Yet repeating the removal of one grain of sand will ultimately leave you with one grain, which is obviously not a heap. Here, it is the system of removal (the MAS), not any of the individual actions in themselves, that turns the heap into a non-heap. Thus, a series of actions that have little or no moral significance when viewed in isolation may amount to a morally impermissible phenomenon when combined.

In fact, even a series of benevolent actions may cause harm when combined, while ill intended actions may amount to something good depending on the constitution of the MAS. Adam Smith's theory of the market economy is a good example; individual actors acting in self-interest result in benefit for society. It is not the sum of the moral significance of actions that matters, but their impact as a MAS. It follows, therefore, that *some* alternatives will seem equally morally permissible considered on the level of individuals but will differ once we consider the MAS of which they are part (see de Font-Reaulx 2017, for a similar argument applied to discrimination).

Moving to towards a solution

Now, let us consider the pervert's dilemma through the lenses of LoA and DM. Much like in the previous examples mentioned, we must approach the dilemma not as the abstracted, hypothetical case of actors A and B , but on a level which takes into consideration the relevant observables (social context) of Deepfakes as a MAS. For by abstracting the Deepfake phenomenon into a matter of “ A ” and “ B ”, one also subtracts from it the very thing that gives it its ethical

significance, namely its role in the social system of gender oppression. In one sentence, you cannot take gender out of pornography, and you cannot take society out of gender. As a *societal phenomenon*, Deepfakes are arguably enabled by a MAS of male consumers, producers, technology, and misogyny. Moreover, it arguably plays a role in the machinery which systematically reduces women (as a collective identity) to sexual objects, even if none of the individual instances can be held to cause this. So it should be fair to say that the phenomenon is *highly* gendered (indeed, one need not spend much time on one of the forums or websites devoted to Deepfakes to realise this). While each isolated video may not affect the women it stars as individuals, the *phenomenon as such*—the MAS—is, in its current form, inseparable from the systematic degrading of women as a collective identity (Dines et al. 1998).

This is why it seems more morally impermissible to use a Deepfake application to create a pornographic video of actress Jennifer Lawrence than of, say, George W. Bush (assuming conditions i and ii as defined above)—even if both are produced for the purpose of sexual pleasure. It is true that both individuals have interests in not having the film made. But when understood through the macro lens of gender inequality—e.g. the technology, the producer, and Bush and Lawrence as parts of a larger system, as opposed to merely two arbitrary individuals—these interests differ in legitimacy. Arguably, it is not a societal problem that rich powerful men are mocked and scorned. Thus the ethical significance of what seems private, and local, lies in the political and social system in which it takes place.

In contrast to Deepfakes, sexual fantasies are not normally considered a gendered phenomenon,³ and there is no immediately identifiable MAS responsible for their existence. Instead, most people have sexual fantasies about others now and then. This is not to say that sexual fantasies do not play a role in gender inequality. Their content most certainly does. And as such, their content also has an ethical significance, as pointed out by Bartel and Cremaldi (2018) and by Corvino (2002), among others. But the fact that the content of sexual fantasies can be impermissible does not mean that sexual fantasies are impermissible per se (Kershnar 2005; Neu 2012). Whereas the content of sexual fantasies may be morally objectionable, few would argue that their mere

³ Few quantitative studies have investigated this in depth. However, the findings of Fisher et al. (2012) suggest that men think about sex about 19 times per day while the same number for women is 10, a difference which is not statistically larger than thoughts about food or sleep. Moreover, for women, the rated importance of social desirability was correlated with lower reported thoughts about sex and food, but not sleep, which suggests a social desirability bias. In sum, sexual fantasies seem far less gendered than what Deepfakes appear to be, although this is yet to be statistically confirmed.

existence (irrespective of content) is grounded in gender inequality. And this distinguishes them from Deepfakes, in which the impermissibility arises regardless of the type of sexual content it contains. In other words, sexual fantasies are, unlike Deepfake Pornography, not a highly gendered phenomenon, and cannot be attributed to any immediately identifiable MAS.

In sum, if we consider the dilemma on high Levels of Abstraction, I find that we have no good reason to deem the consumption Deepfakes more (or less) impermissible than a sexual fantasy. In the example introduced in the introduction to this essay, where A makes a Deepfake video based on B's face, there is no morally relevant difference to a mere sexual fantasy. However, even if each such case is harmless when considered in isolation, the totality amounts to something more than the sum of these individual cases. In fact, the Deepfake phenomenon is so closely connected to its role in gender equality that even when we consider it in the abstract, our intuitions are still guided by the lower societal LoA. This is why the dilemma arises in the first place; we simply cannot “unthink” the societal level. And perhaps we should not. When it comes to sexual fantasies on the other hand, the societal and the individual level do not differ hugely depending on what LoA we take, at least not with regards to moral permissibility.

The LoA approach thus allows us to formulate at least *one* way out of the pervert's dilemma, in which the individual action of creating a Deepfake video (under conditions i and ii) is as morally permissible as a mere fantasy, while the phenomenon—the MAS as considered on a lower LoA—is to be deemed impermissible. Just like in the Sorites paradox, it is (normally) impossible to identify the specific role of the individual in building the whole. The heap of sand consisting of 100,000 grains of sand is no less a heap than one which consists of 100,001 grains. Likewise, the phenomenon of Deepfake Pornography would still be just as harmful were we to remove one individual video, since it is the phenomenon as such, and not the individual cases which are impermissible. Sexual fantasies, on the other hand, are per se permissible, both as a phenomenon as individual instances.

It should be noted that the ethical significance of Deepfakes and sexual fantasies may differ in more ways than merely the one I have pointed to in this essay. For example, to what degree is it permissible to create a Deepfake starring, say, a spouse, who has explicitly given consent to be fantasised about? How would we categorise hypothetical hybrid technologies which help materialise and bring sharpness and durability to already existing imaginary images and perhaps even dreams? Are not all creative technologies a type of hybrid between the imagined and the real? As I am uncertain as to how the LoA approach would be applied in such cases, I do not engage with these angles in this essay, but encourage other to contribute to the topic.

Possible objections

I can already see two possible objections, or limitations, to my approach. The first one regards intermediary scenarios involving less sophisticated technologies. For example, assume that a man uses pen and paper to draw highly realistic pornographic pictures of attractive women he has seen during the day in order to masturbate to. Let us presume that he fulfils conditions (i) and (ii) above, perhaps by destroying the pictures. Is his behaviour morally impermissible? Or just a bit “creepy” in the sense that it diverges from mainstream sexual behaviour? And how can the scenario be unpacked using LoAs? Just like the case of Deepfakes, I believe the answer to this question must be sought in the cultural role of the *phenomenon* of drawing pornographic images of women one has met. To my knowledge, this is not a common practice used in gender oppression in today's society, but in a hypothetical society, it certainly could be.

A second objection may be phrased thus: *Is this not just merely a more sophisticated way of saying “it depends” or “it is more complicated than that”?* To this I can only respond that, on one level—yes indeed, it is a cheap point to make that reality is more complex than the abstracted thought experiment. But the point I have been trying to make is not that we necessarily need more nuance and complexity. What I have been trying to show is that any ethical analysis, as MacIntyre (1981) puts it, requires a preceding sociology—especially when it comes to societal phenomena. This point is, I believe, analogous to Patridge's (2011, p. 308) argument on racist images: “determining if we should reject an imaginative image then might mean knowing quite a bit about the cultural context in which the image is deployed”. Indeed, what is true for race here appears to be true also for gender.

This is not merely saying that moral judgments “depend on one's perspective or context” but is also to make a recommendation as to what perspective (here referred to as LoA) is relevant. To return to Corvino's example of the “Old South Ball”, the moral permissibility of hiring black (consenting) students to pose as slaves for a ball *depends* on whether one focuses only on level of the individuals involved or if one focuses on the black community as a collective with a certain history. Even though both may be plausible, the latter level is the more appropriate.

So, what I have proposed is indeed a more sophisticated way of saying “it depends”, but—I hope—a useful and illuminating way indeed. It is describing in the language of ethics what sociologists take for granted—the link between the individual and the collective. Actions may seem equally morally permissible when we allow a certain level, but different on other levels.

An outlook for further implementations

From the above analysis, it seems that the method of Levels of Abstraction can be employed to generate at least one possible answer to the pervert's dilemma: that Deepfakes are impermissible when considered as a phenomenon and permissible when considered as isolated cases, whereas sexual fantasies are normally equally permissible on both levels. It is plausible, I believe, that the structure of this response can also be applied to other ethical dilemmas.

One example of a moral dilemma where the method of LoA can be employed is the so called *gamer's dilemma* introduced by Luck (2009). The gamer's dilemma refers to the following paradox:

1. Virtual child pornography is morally impermissible.
2. Virtual murder is morally permissible.
3. There is no relevant difference between virtual child pornography and virtual murder (when it comes to moral permissibility).

When taking place in the real world, both murder and paedophilia can easily be condemned on the basis of their negative consequences for the moral patient, but when taking place in the virtual world, no one is directly harmed in either case. Yet, for most people, the moral intuition to condemn virtual child pornography remains very strong. Although there are several important differences—in the gamer's dilemma, only one activity is sexual in nature, and the medium is the same in both activities—the similarity to the pervert's dilemma is unmistakable.

Applying the method of LoA to the gamer's dilemma is beyond the scope of this essay. However, I still wish to draw attention to how my approach can be used to at least open up a new space for discussion. Since the publication of Luck's original article, there have been several attempts to solve the dilemma (Young 2016; Ali 2015; Bartel 2012). A common trait among the proposed solutions appears to be the addition of qualifiers (adding context, to the cases). Thus, the current responses to the gamer's dilemma are, at least to some degree, already lowering the LoA to get to their solution, although none does so systematically. If it is true that the ethical dimensions of an action change with the LoA at which it is considered, it seems that a proper ethical analysis of the gamer's dilemma would also require a full analysis of the social systems in which the actions—virtual murder and virtual paedophilia—take place. Perhaps the difficulty of unpacking the ethical dimensions of the gamer's dilemma therefore stems from the absence of such an analysis.

As mentioned, the LoA approach cannot be fully applied to produce an answer to the gamer's dilemma

within the frames of this essay, but it may provide a more formalised method for locating the gamer's dilemma within a context (or better, at a LoA) where an ethically relevant distinction arises. My method suggests that we should look for a type of solution that acknowledges the ethical insignificance of the isolated action of virtual paedophilia, while at the same time identifying its low LoA significance. Perhaps virtual paedophilia, like the Old South Ball mentioned above, fails to produce individual harm but can be deemed harmful as a phenomenon. The method of LoA allows such a solution to be logically coherent. Moreover, it is plausible that it can be employed not only to the gamer's dilemma but to similar dilemmas in general.

Conclusion

In this essay, I have introduced a new moral dilemma, induced by the emergence of Deepfake Pornography, which I refer to as the pervert's dilemma. My analysis suggests that when the pervert's dilemma is considered on a high LoA—i.e., as isolated cases unrelated to other processes in society—there is no reason why Deepfakes should be deemed more morally impermissible than sexual fantasies. However, when the dilemma is considered on a low LoA—i.e., when we consider the truly morally relevant information—the Deepfake phenomenon can be considered morally impermissible on the basis of its role in gender inequality. The consumption of Deepfakes is undeniably a highly gendered phenomenon, and arguably plays a role in the social degradation of women in society. Sexual fantasies are not.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ali, R. (2015). A new solution to the gamer's dilemma. *Ethics and Information Technology*, 17, 267–274. <https://doi.org/10.1007/s10676-015-9381-x>.
- Bartel, C. (2012). Resolving the gamer's dilemma. *Ethics and Information Technology*, 14, 11–16. <https://doi.org/10.1007/s10676-011-9280-8>.
- Bartel, C., & Cremaldi, A. (2018). 'It' s just a story': Pornography, desire, and the ethics of fictive imagining. *British Journal of Aesthetics*, 58(1), 37–50. <https://doi.org/10.1093/aesthj/ayx031>.
- Cole, S. (2018). We are truly fucked: Everyone is making AI-generated fake porn now. *Vice Motherboard*. Retrieved December 18, 2019 from https://motherboard.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley.

- Cooke, B. (2014). Ethics and fictive imagining. *Journal of Aesthetics and Art Criticism*, 72, 317–327.
- Corvino, J. (2002). Naughty fantasies. *Southwest Philos Review*, 18(1), 213–220.
- de Font-Reaulx, Paul. (2017). What makes discrimination wrong? *Journal of Practical Ethics*, 5(2), 105–113.
- Dines, G., Jensen, R., & Russo, A. (1998). *Pornography: The production and consumption of inequality*. London: Routledge.
- Fisher, T. D., Moore, Z. T., & Pittenger, M. (2012). Sex on the brain? An examination of frequency of sexual cognition as a function of gender, erotophilia, and social desirability. *Journal of Sex Research*, 49(1), 69–77. <https://doi.org/10.1018/00224499.2011.565429>.
- Floridi, L. (2008). The method of levels of abstraction. *Minds and Machines*, 18, 303–329. <https://doi.org/10.1007/s11023-008-9113-7>.
- Floridi, L. (2011). *The philosophy of information*. Oxford: Oxford University Press.
- Floridi, L. (2012). distributed morality in an information society. *Science and Engineering Ethics*, 19, 727–743. <https://doi.org/10.1007/s11948-012-9413-4>.
- Floridi, L. (2013). *The ethics of information*. Oxford: Oxford University Press.
- Harris, D. (2019). Deepfakes: False pornography is here and the law cannot protect you. *Duke Law & Technology Review*, 17, 99.
- Harwell, D. (2018). Fake-porn videos are being weaponized to harass and humiliate women: ‘Everybody is a potential target’. *Washington Post*. Retrieved December 18, 2019 from https://www.washingtonpost.com/technology/2018/12/30/fake-porn-videos-are-being-weaponized-harass-humiliate-women-everybody-is-potential-target/?utm_term=.4adcb8ad9ad2.
- Kershner, S. (2005). The moral status of sexual fantasies. *Public Affairs Quarterly*, 19(4), 301–315.
- Luck, M. (2009). The gamer’s dilemma: An analysis of the arguments for the moral distinction between virtual murder and virtual paedophilia. *Ethics and Information Technology*, 11, 31–36. <https://doi.org/10.1007/s10676-008-9168-4>.
- MacIntyre, A. C. (1981). *After virtue: A study in moral theory*. Notre Dame: University of Notre Dame Press.
- Minton, T. (2017). 12 Dead Celebrities Who Were Resurrected With GCI. Screenrant.com. Retrieved December 18, 2019 from <https://screenrant.com/dead-celebrities-actors-cgi-resurrected-movies-tv/>.
- Neu, J. (2012). *On loving our enemies: Essays in moral psychology*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof>.
- Patridge, S. (2011). The incorrigible social meaning of video game imagery. *Ethics and Information Technology*, 13, 303–312. <https://doi.org/10.1007/s10676-010-9250-6>.
- Smuts, A. (2016). The ethics of imagination and fantasy. In A. Kind (Ed.), *Routledge handbook of philosophy of imagination*. New York: Routledge.
- Young, G. (2016). *Resolving the gamer’s dilemma: Examining the moral and psychological differences between virtual murder and virtual paedophilia*. London: Palgrave Macmillan.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.