Introducing the Pervert's Dilemma: A Contribution to the Critique of Deepfake Pornography

Preprint · August 2019

DOI: 10.13140/RG.2.2.15916.41603

1 author:

Carl Ohman
University of Oxford
5 PUBLICATIONS **30** CITATIONS

SEE PROFILE

Preprint submitted to Ethics and Information Technology February 2019 (currently under review).

Introducing the Pervert's Dilemma: A Contribution to the Critique of Deepfake Pornography

Authors

Carl Öhman[1]

[1]Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford, OX1 3JS, United Kingdom

Corresponding author: carl.ohman@oii.ox.ac.uk

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 an action appears to depend on the degree to which it is abstracted from its 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, 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*, or just "Deepfakes" in daily speech.

Deepfakes 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 pornographic film. 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 an 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. It is thus 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. Although formally banned from sites such as Pornhub, Reddit, and Twitter, a number of new sites have emerged that are specially devoted to sharing, creating, and teaching users how to make their own Deepfakes. 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 can work 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 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.

There is 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. Although Deepfakes strike most people as intuitively disturbing – note that several sites (e.g., Reddit, Pornhub, etc.) have pre-emptively banned such content – it seems difficult to justify this intuition without simultaneously disapproving of other actions not normally considered harmful. For instance, compare Deepfakes to sexual fantasies. Although certain fantasies can be deemed impermissible due to the grotesque nature

of their content (more on this below), the act of fantasising about others in a sexual manner is normally not considered unethical. Yet, 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. Whereas a Deepfake takes a more material form than a fantasy (the materiality of a fantasy can be debated), it is hard to see how this 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 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. 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: *B* has some way of guaranteeing (perhaps by destroying the device) 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 obvious difference in this case to a mere vivid sexual fantasy. Herein lies the dilemma,[1] which can now be can be fully articulated thus:

1. Creating Deepfake videos based on someone's face (without their explicit consent) is morally impermissible.
2. Having private sexual fantasies about someone is (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

[1] 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.

immune to moral criticism (Cooke, 2012), 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 (Kersnar, 2005).

We shall 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 can never be impermissible, while Deepfakes are always are impermissible, but rather that a representation that would (normally) be deemed permissible as a fantasy is *im*permissible as a Deepfake, despite the absence of any obvious morally relevant distinction between the two formats.

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). 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.

When phrased as above, the dilemma is strikingly similar to another moral problem introduced by Luck (2009), known as the *gamer's dilemma*. The gamer's dilemma refers to the following seemingly inconsistent intuitions:

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. To narrow the discussion, Luck (2009, p. 32) makes clear that he limits his scope to cases in which "the character that is virtually molested is controlled by the

computer, rather than another player; the character that is molested clearly represents a child; the game player is an adult; and the game player's character clearly represents an adult." With these conditions in place, it is very hard to identify a morally significant difference between virtual child pornography and virtual murder. Yet, for most people, the moral intuition 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.

Given the vivid literature devoted to solve the gamer's dilemma (Bartel, 2012; Ali, 2015; Young, 2016), interrogating these proposed solutions should constitute a promising starting point for an inquiry into the pervert's dilemma. Even if none of the proposed solutions to the gamer's dilemma can fully solve the pervert's dilemma, they may still function as a roadmap for our strategies. In what follows, I thus go through the most notable contributions to the literature on the gamer's dilemma and examine their applicability to the dilemma of the Pervert. Upon doing so, I shall develop my own approach based on a synthesis of these proposed solutions, which I use to unpack the pervert's dilemma. My approach uses a conceptual method called Levels of Abstraction (Floridi, 2008), inspired by Formal Methods in Computer Science. I will show how this method can be employed to produce at least *one* possible response to the pervert's dilemma. However, I only attempt to use the literature on the gamer's dilemma as a proxy, but make no claims as to solve both problems.

Proposed Solutions to the gamer's dilemma

Luck (2009) lists five possible approaches to resolving the gamer's dilemma, each identifying a unique moral distinction between virtual child pornography and virtual murder. Although none of them fully solves the dilemma, it can be argued each of the subsequent responses in the debate falls into one of these five categories. The first one is based on *social acceptability* – virtual child pornography is simply not as socially acceptable in our culture as virtual murder. While this may surely explain our intuitions, Luck rightly points out that it cannot be used as a *justification*. The second line of argument is that consuming virtual child pornography significantly increases the *likelihood* of a person committing such act in real life, whereas virtual murder does not. This argument is rejected due to lack of empirical evidence, as there is simply nothing suggesting that virtual child pornography leads to real-life crimes. Moreover, it is hard to see how a controlled experiment could establish such a fact in any ethical manner since it would necessitate molestation of real children.

The third type of argument holds that virtual murder is not something gamers do for *pleasure*, whereas child pornography is. Luck rejects this claim as empirically false and names multiple

examples of games where the killing of innocents is a feature frequently used for joy (such as Grand Theft Auto). The fourth possible argument he raises is perhaps the one with the most merit. It claims that virtual child pornography *unfairly singles out a group* for harm. We would for example not accept virtual murder if it only targeted, say, Jews or homosexuals. Yet randomness in victims does not seem to redeem anything in the case of virtual child pornography. We would probably object to a game in which the avatar molested other people regardless of their age and gender, thus including children. Such game would not single out a special group but may still be considered unethical. The fifth type of argument is based on the *special status of children* as particularly innocent and worthy of care. However, it is hard to see how, on this basis, it would be worse to virtually molest a child than to kill an adult. None of the five possible arguments thus seems satisfactory.

Bartel (2012) was one of the first to respond to the dilemma. His approach identifies a type of harm in child pornography – the societal prevention of equality – which applies regardless of whether it is virtual or real. And this type of harm is not prevalent in the case of virtual murder. The argument he mobilises to support this claim is based on Levy's (2002) claim that child pornography (virtual and real) is bad for women, since it eroticizes inequality, thus hindering women from obtaining true equality with men. This means that virtual child pornography is bad since it has negative effects for women's emancipation. The argument can thus be summarized in the following three points:

> (a) that virtual paedophilia amounts to child pornography as it necessarily involves the depiction of sexual acts involving children; (b) that virtual paedophilia is morally objectionable insofar as child pornography is morally objectionable; and (c) that virtual murder is distinct from virtual paedophilia as the latter necessarily involves child pornography while the former does not.

Bartel's argument can be challenged on several grounds. Young (2016, p. 67) for instance, questions the claim that virtual paedophilia is child pornography. Moreover, while there is no logical fault with the argument (if virtual child pornography *is* child pornography, and child pornography is bad for women, virtual child pornography is morally impermissible), one may question whether the eroticization of inequality really prevents true equality. Bartel seems to base the argument on an empirical hypothesis – that women are de facto harmed by eroticization of inequality – and there is little research to support this hypothesis. Besides, if eroticization of inequality is morally impermissible, this would disqualify a significant portion of "mainstream" sexuality. As Stephen-Davidowitz shows in his book *Everybody Lies* (2016, p. 93), about 25% of all searches on Pornhub made by female users are in some way related to power asymmetries, violence, or humiliation. It therefore seems questionable to be too judgemental towards those who enjoy eroticization of inequality.

The standard approach to the problem (as proposed by Luck and Bartel) is focused on finding a difference between virtual pornography and virtual murder. In contrast to such approaches, Ali (2015) proposes that we should instead deny the very premises that the Gamer necessarily finds *all* instances of virtual child pornography impermissible and *all* instances of virtual murder impermissible. Ali argues that attitudes which seem clear-cut and homogeneous (e.g., virtual murder is permissible; virtual child pornography is impermissible) are in fact rather diverse. Some instances of virtual murder *are* actually found morally impermissible by most people's intuition and some instances of virtual child pornography are actually not. This depends on the in-game context according to Ali. Some enactments that we would normally feel are intuitively impermissible when in the context of simulation games[2] suddenly seem fine when merely a feature in a sporting game, i.e. just a means to another end. Thus, Ali's strategy is to try to dissolve, rather than resolve the problem. While Luck (2018) acknowledges that Ali dissolves a "weak" version of the dilemma, he points out that one may still imagine cases in which the problem certainly remains. For example, a game played for sport that involves molesting as many children as possible does not seem intuitively permissible, even if it is just a means to another end. Moreover, the distinction between games that are designed as competition, narrative, and simulation seems irrelevant to the pervert's dilemma; the fact that the Deepfake video is only part of a game seems to add little redemption.

Finally, we must consider Young's (2016) argument. Young uses a framework that he calls Constructive Ecumenical Expressivism (CEE), which aims not to identify a morally significant difference between virtual murder and virtual child pornography, but rather to interrogate the relationship between our attitudes (the intuition) and the object in question. Young argues that because virtual enactments lack a natural context, they can be interpreted in many different ways. This enables a pluralism of (independent) beliefs about why paedophilia is morally impermissible. Put together, these reasons form a norm. From the meta-ethics of CEE, Young thus develops a normative stance which holds that a shared moral attitude of a given society (i.e., a norm) constitutes a good justification for deeming an object (an act or phenomenon) morally impermissible, insofar as that attitude is comprised of a large set of *de re* attitudes against that object. That is to say that the larger the amount of justified de re attitudes against something, the stronger the reason to reject it, regardless of the *de dicto* reasons for rejecting it.

So, to put it in Young's (2016, p. 123) own terms: "the premise on which the dilemma is built stems from a difference in attitude which itself is not based on a single factor or a single morally relevant difference. To resolve the dilemma, one would need to undermine each or a large number of the

[2] Ali differentiates between three different kind of games: (1) games played as sport, (2) games played as a narrative and (3) simulation games. Depending on what kind of game is played, the moral implications of an action differ.

different beliefs which ground the single moral attitude (objectified social norm).” The reason this is so difficult is because each belief is based on a certain interpretation of the virtual enactment.

Although it has some merits, it appears that this approach runs into the problem of what defines a “justified” moral belief. To Young, it seems that the quantity of justified beliefs itself amounts to a normative weight. But, as Dennison (2016, p. 238) points out, “Young provides no justification as to why the reader is to accept his move from the descriptive to the normative”. Surely, there are numerous historical examples of societies with moral beliefs that we would consider abhorrent today. Besides, who is to determine what is justified? And according to what principles? Young himself gives sparse consideration to these fundamental questions.

Although none of the proposed solutions seems to fully succeed in resolving the dilemma, they do nevertheless all have some merit, and thus suggest certain strategies that may be applied to solve the pervert’s dilemma. A common trait among the proposed solutions appears to be the addition of qualifiers. That is, they go beyond the original problem as laid out by Luck and highlight external conditions on which our moral judgment may depend. For example, Bartel (2012) seeks the solution to the dilemma on the societal level, as opposed to the analytical. And Ali explicitly removes the problem from its abstracted form and argues that the answer depends on the *in-game context*. The permissibility of the actions in question can only be evaluated in relation to their immediate environment.

Young goes for a similar strategy in pointing to the importance of the gamer’s “interpretation” and the fact that virtual enactments allow so many of them. This notion of “interpretation” can also be seen as a form of contextualisation – it is the player’s attempt to connect what she sees to *something else* beyond the frame of the actual game. The information required to solve the dilemma is therefore located not just in the context of the game, but in its relation to the context created by the player, i.e. its relationship to the external environment. The permissibility of the actions depends on the context in which they occur.

Unfortunately, “it depends” it not a very satisfying answer. When someone asks us to justify our moral judgements, answering that “it depends on the context” is about as helpful as a shrug. Although it is quite possibly true that our moral judgement must depend on the context in which the action in question takes place, a moral analysis must be more specific. We must identify precisely *how* it depends, upon *what* it hinges. In other words, we need to formalise “it depends”. In the following section, I shall outline how this can be achieved by employing the method of Levels of Abstraction. Moreover, I shall argue that the totality of a series of actions with negligible moral significance may

amount to something beyond the sum of its parts. To be clear, this is still saying "it depends", only saying more precisely *how* it does so.

The Method of Levels of Abstraction: Formalising "It Depends"

Both the Gamer's and the pervert's dilemma are induced by the emergence of sophisticated information technology. For this reason, it makes sense that our approach to unpacking the problems should also take an informational viewpoint. That is, we should understand *A*, *B*, and their actions as agents who act in response to some kind of informational environment. From this perspective, the question to ask becomes: What 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 Levels of Abstractions 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 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. An 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 (e.g. alternative moral actions), be it technologically, morally 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 endless 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 that the method of LoAs is in no way a relativist approach. A question is always asked for a purpose – in order to attain some form of information – and for that specific purpose, there is an appropriate LoA. 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 19th 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 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), (on a very high LoA, that is). 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 a case of murder and not manslaughter or mere self-defence, we need to know very little 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” (what I refer to as 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, the second type of harm may turn out to be prevalent even if it should turn out the former was absent. 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 weIl was beside the point.

What Corvino describes here is a clash between two LoAs – one which focuses on the individuals involved and one which focuses on the collective identity as the object of harm. Both sides are right, but the latter level is arguably more relevant because it engages ethically important observables. The lower LoA here contains what Patridge (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. 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. Yet the latter only arises when we consider 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 ∨ 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 in ‘‘supranational’’) C = A∪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 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 in itself (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 1 grain, which is obviously not a heap. Here, it is the system of removal, not any of the individual actions in themselves, that turn the heap into a non-heap. Thus, a series of actions that have little or no moral significance when viewed in isolation may when combined amount to a morally impermissible phenomenon.

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 example, we must approach the dilemma not as the abstracted, hypothetical case of actors *A* and *B*, but on a MAS level which takes into consideration the social context (relevant observables) of Deepfakes. 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. 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, Donald Trump (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 Trump 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 obvious MAS responsible for their existence. Instead, most people have sexual fantasies about others now and then, and it is normally not something one would object to being a target of. 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 & 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.* Whereas the content of sexual fantasies may be morally objectionable, few would argue that their mere existence (regardless of content) is grounded in gender inequality. And this distinguishes them from Deepfakes, in which the impermissibility arises regardless. In other words, sexual fantasies are, unlike Deepfake pornography, not a highly gendered phenomenon, and cannot be attributed to any obvious MAS.

In sum, if we consider the dilemma on high Levels of Abstraction, I find that we have no good reason to deem Deepfakes more (or less) impermissible than a sexual fantasy. However, even if each individual case is harmless when considered in isolation, the totality amounts to something more than

[3] Few quantitative studies have investigated this in depth. However, the findings Fisher et al. (2012) suggest that men think about sex about 19 times per day while the same number for women is 10, which is not statistically larger than thought 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.

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.

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 pornographic pictures of attractive women he has seen during the day. 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"? 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".

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 other ethical dilemmas – such as the gamer's dilemma – are also dependent on the LoA at which they are analysed. However, as pointed out in the introduction to this essay, there are some significant differences between the Pervert's and the gamer's dilemmas. And for this reason, applying the method of LoA to the latter is beyond the scope of this essay. Instead, I wish to draw attention to how my approach can be used to at least open up a new space for discussion.

The aforementioned responses to the gamer's dilemma that guided my solution to the pervert's dilemma, are, at least to some degree, already lowering the LoA to get to their problem, although none does so explicitly. What I have been arguing in this essay, is largely that the ethical significance of an action depends on the action's position in the social system within which it takes place. Even a selfish act can have positive outcomes if properly and systematically related to other selfish actions (e.g. the market economy). Similarly, the ethical significance of actions must sometimes be considered at the level of the social system in which they are situated. For some problems, this task is rather straight forward. The ethical significance of the pervert's dilemma must be sought at the societal level of gender oppression, and the ethical significance of Corvino's example of the Old South Ball must be sought in the history of slavery and racial oppression. It is less clear, however, where exactly the solution to the gamer's dilemma should be sought.

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 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 stems from the absence of such an analysis. Among the responses to the gamer's dilemma, only Bartel (2012) engages the lower LoA (i.e., discusses the societal dimensions of the phenomenon). Although the argument proposed by Young (2016) also engages the societal level, it fails to do so in a way that properly develops a normative stance. I claim that the LoA approach may resolve these problems.

The LoA approach cannot be fully applied to produce an answer to the gamer's dilemma within the frames of this essay, but it does sketch out what type of solution we should be looking for. It 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. The method of LoA allows such a solution to be logically coherent.

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. Because of its striking resemblance to the gamer's dilemma, I have analysed the proposed solutions to the latter and used them to form a way out of the pervert's dilemma – the method of Levels of Abstraction and Distributed Morality.

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.

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

Cooke, B. (2014). Ethics and Fictive Imagining. Journal of Aesthetics and Art Criticism 72, 317–327.

Cole. S. (2018). We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now. Vice Motherboard. Retrieved from: https://motherboard.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley

Corvino, J. (2002). Naughty fantasies. Southwest Philosophy Review, 18(1).

Dennison, R. (2017). Gary Young, Resolving the gamer's dilemma: examining the moral and psychological differences between virtual murder and virtual paedophilia. Ethics Inf Technol. 19(3): 237. https://doi.org/10.1007/s10676-017-9434-4

Dines, G., Jensen R. & Russo, A. (1998). *Pornography: The Production and Consumption of Inequality*. London: Routledge

Ellis B.J., & Symons D. (1990). Sex differences in sexual fantasy: An evolutionary psychological

approach. *Journal of Sex Research*, 27:4, 527-555. DOI: 10.1080/00224499009551579
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,* DOI:10.1018/00224499.2011.565429
de Font-Reaulx, Paul (2017). What Makes Discrimination Wrong? *Journal of Practical Ethics* 5 (2):105-113.
Floridi, L. (2008). The Method of Levels of Abstraction. Minds & 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**.** *17 Duke L. & Tech. Rev. 99*
Harwell, D. (2018). Fake-porn videos are being weaponized to harass and humiliate women: 'Everybody is a potential target'. *Washington Post*. Retrieved 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
Kershnar, S. (2005). The Moral Status of Sexual Fantasies. Public Affairs Quarterly, 19(4), 301–315.
Luck, M. (2018). Has Ali dissolved the gamer's dilemma? Ethics and Information Technology (2018) 20:157–162. https://doi.org/10.1007/s10676-018-9455-7
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 DOI 10.1007/s10676-008-9168-4
Minton, T. (2017). 12 Dead Celebrities Who Were Resurrected With GCI. Screenrant.com. Retrieved from: https://screenrant.com/dead-celebrities-actors-cgi-resurrected-movies-tv/
MacIntyre, A. C. (1984). After virtue: A study in moral theory. Notre Dame, Ind: University of Notre Dame Press.
Neu, J. (2012). On Loving Our Enemies : Essays in Moral Psychology The Ethics of Fantasy. In On Loving Our Enemies: Essays in Moral Psychology. Published to Oxford Scholarship Online: https://doi.org/10.1093/acprof
Patridge, S. (2011). The incorrigible social meaning of video game imagery. Ethics Inf Technol. 13: 303–312. DOI 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.
Stephen-Davidowitz, S. (2016). Everybody lies : big data, new data, and what the Internet can tell us about who we really are. New York: William Morrow & Co
Young, G. (2016). *Resolving the gamer's dilemma: Examining the Moral and Psychological Differences Between Virtual Murder and Virtual Paedophilia*. Palgrave Macmillan: 10.1007/978-3-319-46595-1