

**Title:** “Deepfake Technology and Individual Rights”

**Abstract:** Deepfake technology can be used to produce videos of real individuals, saying and doing things that they never in fact said or did, that appear highly authentic. Having accepted the premise that Deepfake content can constitute a legitimate form of expression, it is not immediately clear where the rights of content producers and distributors end, and where the rights of individuals whose likenesses are used in this content begin. This paper explores the question of whether it can be plausibly argued that Deepfake content involving the likenesses of real individuals violates the rights of these individuals.

**Keywords:** deepfake technology, individual rights, privacy, defamation, property, free expression

**Publication information:** This is an upcoming paper for the journal *Social Theory and Practice*. Copyright for the paper is held by *Social Theory and Practice*. Information about this journal is available here: <https://pdcnet.org/soctheory>.

## (I) Introduction

“Deepfake” technology employs Artificial Intelligence in order to manipulate video and audio content in a manner that is undetectable to consumers of this media. Deepfake technology can be used to produce videos of real individuals, saying and doing things that they never in fact said or did, that appear highly authentic. Controversy is already emerging with regard to this technology, as pornographic Deepfake content is being produced that exploits the likenesses of celebrities without their consent. Having accepted the premise that production and distribution of Deepfake content can constitute a legitimate form of expression, it is not immediately clear where the rights of content producers and distributors end, and where the rights of individuals whose likenesses are used in this content begin. This paper explores the question of whether it can be plausibly argued that Deepfake content involving the likenesses of real individuals violates the rights of these individuals. Three arguments in support of the view that Deepfake content can be legitimately constrained, on the grounds that it infringes upon individual rights, are considered. I name these the Privacy Argument, the Defamation Argument, and the Property Argument.<sup>1</sup> My view is that the Property Argument provides the most robust and reliable strategy for assessing the legitimacy of Deepfake content and determining whether legal constraints should be placed on it. I argue that the strength of this argument lies in its relative simplicity and parsimony.<sup>2</sup> I then offer a brief, supplementary discussion of “marketplace solutions,” i.e. strategies for regulating the

---

<sup>1</sup> This is not a policy paper. While there are no doubt many laws in place in various jurisdictions that are related to privacy, defamation, and property, throughout this discussion I avoid examining any specific law in any specific jurisdiction. Since Deepfake technology has not yet been analyzed through a philosophical lens, the aim of this paper is to lay a conceptual groundwork that can shape future discussions about Deepfake technology and individual rights with a more direct emphasis on policy.

<sup>2</sup> A philosophical literature is currently being developed that explores the relationship between online data collection and individual property rights. I believe that while there is room for intersection and overlap between this area and my own Property Argument, which is concerned specifically with Deepfake technology and the appropriation of individuals' likenesses, it is interesting to consider whether this argument ought to be situated within a broader debate about data ownership and collection. I will let readers arrive at their own conclusions with regard to this matter, as I here specifically emphasize issues surrounding Deepfake technology and do not deviate into the broader debate about data. Authors who have explored this area include Janeček (Janeček 2018) and Vold & Whittlestone (Vold & Whittlestone 2019).

production and distribution of Deepfake content that do not (necessarily) need to appeal to the concept of individual rights in order to retain their plausibility.

The arguments presented in this paper emphasize the language of individual rights rather than the language of harm and harm prevention. It is important to identify the theory of rights that informs this discussion, as well as its relationship with the concept of harm. The arguments that follow operate within a broadly liberal framework, in the sense that they accept as a premise that the state can only legitimately constrain other-regarding behaviours (as opposed to self-regarding behaviours) that cause harm. The fact that a behaviour is widely or strongly disliked is insufficient with respect to justifying state intervention seeking to constrain said behaviour. Within this liberal framework, to charge one with a rights violation also entails charging one with infliction of harm. To invade the privacy of an individual is thus to inflict harm upon that individual, and the same is true of defaming an individual, as well as illegitimately seizing an individual's personal property.

This liberal framework employs a Millian approach to limiting harm through force that is more demanding than alternative approaches.<sup>3</sup> We can easily understand that making a cruel comment towards someone or being unfaithful towards a romantic partner can be harmful behaviours in the sense that they inflict pain that ought to be avoided, yet these behaviours do not satisfy the conditions required in order to justify state intervention within a Millian approach. While there are many possible deployments of Deepfake technology that are capable of having adverse effects on individuals, this fact alone does not establish that it is legitimate for such content to be constrained through force. Mill's Harm Principle<sup>4</sup> underpins a set of negative political rights,

---

<sup>3</sup> Of course, there may be good reasons to doubt whether Mill's understanding of harm is sound. I do not open that debate here. My arguments about Deepfake technology and individual rights are downstream from the debate about how harm ought to be understood.

<sup>4</sup> John Stuart Mill's harm principle, articulated in *On Liberty*, famously holds that "[t]he only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others" (Mill 2015: 13). Sumner explains that "[u]nder the Harm Principle harm to others is a necessary, but not a sufficient, condition for social interference with an activity" (Sumner 2004: 25).

and the focus of this discussion is whether certain deployments of Deepfake technology can be determined to be at odds with (some of) these rights. Concerns about harm thus provide a conceptual backdrop for this discussion about individual rights, despite the fact that the arguments presented emphasize the language of rights rather than directly invoking the language of harm.

## (II) The Privacy Argument

There are many philosophical views regarding the importance of privacy. One view is that privacy is foundational to freedom more generally. Spinello argues that,

there is an especially close relationship between privacy and freedom. It is quite difficult to exercise the liberties guaranteed by the Constitution when our actions are on display or when much of the intimate information about our lives is in the public domain. It is a struggle to make authentic choices when our evaluations, preferences, past history, and future objectives become widely known without our consent. (Spinello 1998: 727)

Others highlight the significance of privacy as a mechanism for cultivating functional social relationships. In *Privacy and Social Freedom*, Schoeman highlights the importance of privacy as a mechanism for cultivating “associations and relational ties” between people (Schoeman 1992).<sup>5</sup>

The relationship between technological advancement and privacy has often been a fraught one.<sup>6</sup> It seems that as the scope of technological capability increases, the scope of what can be kept

---

<sup>5</sup> While this view of the value of privacy seems sound, it seems to me that it would be difficult or impossible to formulate a view of privacy *rights* that is animated by the priority of preserving and fostering valuable relationships. In the case of pornographic Deepfake content, it is understandable why one would argue that this content has harmed their relationships, and that it is therefore morally problematic. However, I do not see how one can extend this line of reasoning to the conclusion that their privacy rights have therefore been violated.

<sup>6</sup> Vold & Whittlestone argue that “because it is becoming easier for companies to use collected data for influence, threats to privacy are increasingly also threats to personal autonomy—an individual’s ability to reflect on and decide freely about their values, actions, and behaviour, and to act on those choices” (Vold & Whittlestone 2019: 3). We might accordingly say that privacy and autonomy are mutually reinforcing, and that technologization has the potential to erode both simultaneously.

private decreases.<sup>7</sup> Consider the ubiquity of mobile applications such as Google Maps that enable application developers to record the movements of smartphone users through physical space.<sup>8</sup> While reasonable observers may disagree about whether these types of applications infringe users' privacy rights, there is virtually no disagreement about the fact that the decision to use such applications does indeed result in diminished privacy.<sup>9</sup> As services such as Google Maps have become staples of everyday life for an increasingly large segment of the world's population, discussions about privacy have tended to focus on whether the loss of privacy that the services entail is problematic (and if so, just how problematic) rather than whether privacy is indeed being lost.<sup>10</sup> The notion that using these services involves some sacrifice with regard to personal privacy is straightforward and uncontroversial.

This is not the case for Deepfake technology. Deepfake technology stands in contrast to other technologies such as web mapping services, as one may plausibly argue that the appropriation of a person's likeness for the production of Deepfake content involves no loss of privacy for the targeted person.<sup>11</sup> Assuming that the person's likeness has already been publicized through some other means, Deepfake content producers need not reach into the private life of their target in order to retrieve data as a means of producing a final product that appears highly authentic. Thus the thought that Deepfake content does not involve a loss of privacy for those whose likenesses are used intuitively seems sound. Of course, if producers of Deepfake content are

---

<sup>7</sup> Anderson points out that “[v]irtually all of the significant developments in the modern theory of privacy in the U.S. have in their very near vicinity a technological or social-organizational development to which they respond” (Anderson 2008: 101-102). Spinello warns that “If our privacy continues to evanesce in the wake of technology's unrelenting progress so too will our basic freedoms” (Spinello 1998: 727).

<sup>8</sup> An investigation by the Associated Press produced evidence that Google can and does continue recording location information about users even after they have turned off the “Location History” setting on the Google Maps app (Nakashima 2018).

<sup>9</sup> Douglas notes that “[f]requently updated location information, such as stored by mobile devices that record their location, may reveal an individual's daily routine, and so establish pattern knowledge about that individual” (Douglas 2016: 201).

<sup>10</sup> My view is that having one's location data collected constitutes a loss of privacy even if no human is examining this data directly.

<sup>11</sup> In an 1890 *Harvard Law Review* article, Warren and Brandeis famously summarized the right to privacy as “the right to be let alone”. If this conceptualization of privacy is accepted, then producers of Deepfake content can plausibly argue that their work is consistent with “letting alone” the individuals whose likenesses are used, and thus has no impact on the privacy rights of these individuals. Of course, other conceptualizations of privacy are possible.

committing acts such as trespassing, stalking, or hacking in order to achieve their final product, this will clearly constitute a breach of the right to privacy. But the fact that Deepfake content production can theoretically take place in tandem with blatantly immoral and illegal acts cannot be used to demonstrate the illegitimacy of Deepfake content production itself. We must consider whether Deepfake content production can constitute an invasion of privacy in the absence of unambiguously invasive actions such as those mentioned above.<sup>12</sup>

A plausible argument for the view that Deepfake content production can constitute an invasion of privacy may appeal to the principle of personal impact.<sup>13</sup> If Deepfake technology can be used to produce video content depicting an individual in a private situation that they would never consent to sharing with the public, and this video product is indistinguishable from an authentic video product, then the individual targeted by the Deepfake content may argue that the impact on them is also indistinguishable from an “authentic” act of privacy invasion. It is already the case that Deepfake technology is being used to produce pornographic content that exploits the likenesses of celebrities without their consent. Since the targeted celebrity did not agree to being filmed in sexually explicit scenarios and having this footage distributed to the public, at one point in time the only way for a content producer to obtain and distribute such footage would be to film this celebrity in sexually explicit scenarios without their consent,<sup>14</sup> which obviously constitutes an

---

<sup>12</sup> Many have noted the lack of a widely accepted definition of privacy, as well as the lack of a widely accepted philosophy of privacy rights, with Judith Thomson noting that “[p]erhaps the most striking thing about the right to privacy is that nobody seems to have any very clear idea what it is” (Thomson 1975: 295). Throughout this discussion I set aside the larger philosophical questions surrounding the nature of privacy rights, confining myself to considering how privacy rights can plausibly be appealed to in order to justify placing constraints on Deepfake technology.

<sup>13</sup> I have chosen to use the language of “personal impact” here rather than the more familiar language of “victim impact” because I want to avoid creating the impression that I am begging the question of whether an individual has in fact been victimized in this or that case involving Deepfake technology. If the question of whether victimization has in fact taken place is open to debate, then it is better to use the neutral language of “personal impact” in order to avoid confusion or (ostensible) circularity. (For a treatment of victim impact statements and their relationship to liberalism, see Rosebury 2011.)

<sup>14</sup> Scenarios like this are not merely hypothetical. Terry Gene Bollea, professionally known as “Hulk Hogan”, was awarded \$115,000,000 in compensatory damages after Gawker Media released a sexually explicit video of him that he maintains was recorded without his consent.

invasion of privacy. Virtually all observers would agree that this is an immoral and illegitimate act.

From the perspective of the celebrity, the problem now is that technology exists that enables exploiters of celebrities to produce content that is indistinguishable<sup>15</sup> from the genuine article without having to do the actual “work” of filming the celebrity in sexually explicit scenarios without their consent. The line of reasoning here is that while some steps have been bypassed, so to speak, in the process of exploiting celebrities via invasion of privacy, the actual impact on the targeted celebrity is no different than the impact generated by more conventional forms of privacy invasion. An individual may feel equally victimized by a pornographic Deepfake video of them produced without their consent as they would by an authentic pornographic video of them that had been produced without their consent. Since Deepfake content is (effectively) indistinguishable from authentic content from the perspective of consumers, it should be treated as such from the perspective of individual rights. The fact that the Deepfake content producers did not *technically* violate the privacy of their target is just that - a technicality.<sup>16</sup> In these contexts, if privacy is to be conceptualized in a robust and substantive manner,<sup>17</sup> then reasonable critics are bound to the conclusion that the celebrity’s right to privacy has indeed been violated. This is the Privacy Argument.<sup>18</sup>

---

<sup>15</sup> I do not wish to give the impression that all Deepfake content is highly convincing. However, it seems reasonable to predict that as this technology continues to improve and become more available, the proportion of Deepfake content that is highly convincing will increase. While technology is available that can be used to detect Deepfake content, the fact that Deepfake technology is rapidly evolving means that there is currently no stable, reliable method for detecting Deepfake content. In 2019, it was announced that Facebook would be sponsoring an event known as the Deepfake Detection Challenge in hopes of developing new tools for accurately identifying Deepfake content (Metz 2019).

<sup>16</sup> In a lecture that is available for viewing online, journalist Samantha Cole makes the following statement, which seems to evoke this type of sentiment: “Targets of Deepfakes say that it does not matter intellectually that they know the images are fake. It feels like a violation. It feels like hundreds of people have seen them having sex on the Internet with a stranger. It feels real” (Cole 2020).

<sup>17</sup> It is worth pointing out that the US Supreme Court’s ruling in *Roe v. Wade* is founded on the positing of a right to privacy. In this case, it was determined that a ban on abortions is unconstitutional because it violates the right to privacy. If one finds the language of a “robust and substantive” conceptualization of privacy curious, the *Roe v. Wade* ruling may provide a glimpse of what such a conceptualization looks like in practice.

<sup>18</sup> Of course, this argument is founded on the premise that individuals do indeed possess a right to privacy. Anderson outlines two main ways in which the right to privacy can be conceptualized. It can be conceptualized as protecting private individuals against the state, or it can be conceptualized as protecting private individuals from the state *as well as* other private actors. This discussion is informed by the latter conceptualization (Anderson 2008).

While this argument does have merits, I think that these merits are outweighed by some serious flaws. The biggest flaw relates to the amorphous character of the principle of personal impact. It is absolutely true that Deepfake content can impact targeted individuals in a way that causes them to feel that their privacy has been seriously compromised. But this is also true of other types of content. Many pieces of media can lead to sincere and acute feelings of one's privacy being violated, and yet they are nonetheless viewed as legitimate forms of expression that do not entail a violation of anyone's rights. Using personal impact as a criterion in order to determine whether invasion of privacy has taken place may render many forms of media vulnerable to the charge that they have violated the rights of particular individuals.<sup>19</sup>

If a biopic<sup>20</sup> is produced that is unflattering towards the person being biographed, then the person or their estate may have strong feelings about it that relate to the issue of privacy. Indeed, they might plausibly argue that the loss of privacy that they have experienced is greater than the loss of privacy experienced by those who have had their likenesses used in pornographic Deepfake content. While being the target of a pornographic Deepfake video may feel very invasive, a biopic that reveals intimate details of one's life (related to personal relationships, physical and mental health, professional setbacks, etc.) may feel even more invasive. Indeed, feelings regarding invasiveness will likely vary greatly, as individuals are quite diverse. Every individual is unique

---

<sup>19</sup> There is a significant gap between producing Deepfake content and publishing Deepfake content. One issue that remains open for debate is whether the Privacy Argument can be plausibly aimed at both of these forms of activity, or whether it ought to solely be concerned with Deepfake content that has been published in addition to being produced. A legal case in the Canadian province of Quebec known as *Aubry v Éditions Vice-Versa Inc* arrived at the conclusion that while a photographer does indeed have a right to photograph a person in a public space without consent from the subject, they do not have the right to publish this photograph without consent. It is interesting to consider whether this reasoning can plausibly be applied within the domain of Deepfake content. Granting that it is illegitimate to publish a pornographic Deepfake video depicting a celebrity without consent, is it also illegitimate to produce such content solely for the sake of one's own personal use? This question merits further investigation. Sociologist Gary T. Marx explores this theme involving the scope of the private and the public in his book *Windows into the Soul: Surveillance and Society in an Age of High Technology* (specifically, a chapter titled "The Private within the Public: Psychological Report on Tom I. Voire") (Marx 2016).

<sup>20</sup> The word "biopic" is an abbreviated combination of the words "biographical" and "picture". The films *A Beautiful Mind*, *Ray*, and *Milk* are all examples of the biopic genre.



with regard to their personal boundaries and what they deem to be appropriate for publicization.<sup>21</sup> Some individuals may be horrified at the prospect of being the focus of an unflattering biopic, but be relatively unbothered by the prospect of being the target of a pornographic Deepfake video. Of course, the reverse could also be true.

The upshot is that if we accept the principle of personal impact, and use it to formulate a litmus test that aims to determine whether privacy rights have been violated, the outcomes could be far too inclusive, and lay the groundwork for the censorship of many types of content that are (rightfully, in my view) conceptualized as legitimate forms of expression. While I have chosen the example of biopics for this discussion, we can easily imagine many documentarians, journalists, satirists, comedians, and so on, being accused of privacy violations in a world wherein personal impact is used as the unit of measurement that determines whether a privacy rights violation has occurred. Indeed, a great deal of art is revered precisely because it communicates information and insights that certain individuals (particularly powerful individuals) would prefer to keep quiet. If the Privacy Argument is accepted, we may find that it legitimizes censorship towards a wide variety of media as collateral damage, so to speak.<sup>22</sup>

It is important to emphasize that my conclusion regarding the Privacy Argument and the principle of personal impact in no way dismisses the fact that individuals may be genuinely aggrieved by Deepfake content. To the contrary, the potentially offensive character of Deepfake content is part of what motivates my discussion in the first place. However, individuals living in pluralistic societies are caused offense by innumerable things. Some individuals are offended by interpretations of a religious text that rival their own interpretation, and sometimes these

---

<sup>21</sup> Spinello notes: "...privacy lacks the simplicity of other collective goods or values since it affects so many areas of our lives in different ways and since personal preferences about privacy vary so widely. Some people simply *do* prefer more of it than others" (Spinello 1998: 730).

<sup>22</sup> McCloskey raises a concern that is effectively the mirror image of this one. His concern is that linking the concept of privacy to "emotional reactions" may be underinclusive, as individuals living under totalitarian regimes may be conditioned to tolerate and accept violations of their privacy that ought to be resisted (McCloskey 1980: 28).

differences in interpretation can generate so much animosity that they lead to physical violence. Regardless, it is generally agreed upon by those with a liberal orientation that an interpretation of a religious text cannot legitimately be censored on the grounds that it causes offense to those who disagree with it.<sup>23</sup> I subscribe to the view that proving that an individual has been offended by a form of expression is not equivalent to proving that an individual has been harmed by a form of expression.<sup>24</sup> This view does not preclude the possibility of expression sometimes entailing genuine harm towards individuals. If it turns out to be the case that a certain form of expression, such as a biopic, has functioned to damage the reputation of its subject in a manner that is undeserved and unwarranted, then this will provide grounds for the charge that this biopic has served to violate the rights of its subject. However, this charge brings us outside of the domain of privacy violations and into the domain of defamation. This domain is the focal point of the following section of my discussion.

### **(III) The Defamation Argument**

Another line of argument with regard to the relationship between Deepfake technology and individual rights revolves around the concept of defamation. Defamation is generally well understood as a concept, and most, if not all, proponents of free expression view this concept as legitimate. While it is true that individuals have the right to express themselves freely without fear of (legal) punishment, it is also true that people at times abuse this right in order to damage the reputations and livelihoods of others. When such abuses take place, it is legitimate to punish those

---

<sup>23</sup> Feinberg advances the view that while there is a distinction between harm and offense (with harm being greater in severity), some forms of offense are great enough to warrant prohibition. Feinberg thus offers an “offense principle” to complement the Millian harm principle. While I do not accept Feinberg’s view, it would be interesting to examine the relationship between Deepfake technology and the offense principle (Feinberg 1985).

<sup>24</sup> My position is largely informed by arguments advanced by L.W. Sumner in his book *The Hateful and the Obscene*. Sumner argues in favour of the view that censorship can only be justified (i.e. legitimate) if and when it can be demonstrated that a form of expression produces harm.

guilty of committing acts of defamation; this punishment does not constitute a violation of one's right to free expression.<sup>25</sup> Importantly, when an agent (whether this agent be an individual or a group) is charged with committing an act of defamation, the debate surrounding whether defamation has taken place will generally emphasize the issue of impact rather than the issue of intent. While intent can be taken into consideration when assessing charges of defamation, ultimately one can be guilty of defamation regardless of whether their intent is malicious, noble, or neutral. What is of primary interest is whether the words and ideas propagated by the agent have served to damage the reputation of another in a manner that is illegitimate.<sup>26</sup>

In certain contexts, misbehaviour involving Deepfake content will fit fairly snugly into the conceptual schemes that have been established with regard to defamation, which subsumes the categories of slander and libel. If a media outlet circulates a Deepfake video of a prominent individual committing a heinous crime and claims that the video is authentic, it is plain to see that this constitutes libel, and that the rights of the individual in question have been violated. It is also plausible to advance the supplementary claim that since Deepfake content appears authentic in a manner that is distinct from printed content (or other types of media), it constitutes an especially grave form of libel, and therefore an especially serious rights violation. It seems reasonable to suggest that defaming someone through video and audio that have been manipulated by Artificial Intelligence is more egregious than defaming someone through printed or spoken words, while maintaining that both are unacceptable.

---

<sup>25</sup> Peonidis pithily articulates this point: "It is intuitively evident that in certain contexts defamatory speech can have dire consequences no liberal can condone" (Peonidis 1998: 8).

<sup>26</sup> An implication here is that oftentimes damaging one's reputation is indeed legitimate. A news outlet can cause damage to the reputations (and livelihoods) of many individuals without ever engaging in conduct that is defamatory, so long as the damaging information they publish is correct.

In other contexts, the Defamation Argument may run into problems. Consider again the example of pornographic Deepfake content that exploits the likenesses of celebrities.<sup>27</sup> Does it make good sense to view this sort of content as a form of defamation? There are a handful of ways in which producers of this content can attempt to sidestep the accusation that they have defamed a celebrity. The manner in which the content is presented is significant. Deepfake content can be presented in a manner wherein consumers are explicitly informed that their senses are going to be deceived. The realism of Deepfake content need not be accompanied by the idea that consumers of this content ought to think that it is authentic. Just as producers of video games can use the realism of these games as a selling point, while preserving the clear understanding between producer and consumer that the events taking place on screen<sup>28</sup> are fictitious, a producer of Deepfake content can cultivate a similar understanding. Deepfake content can feature disclaimers that explicitly highlight the lack of authenticity of the events taking place on screen. It may be the case that disclaimers like these actually do alleviate a great deal of concern about defamation. It is plausible to think that targeted individuals, and those who are sympathetic towards them and care about their rights, might be significantly less bothered by pornographic Deepfake content if it includes such disclaimers. Even if they dislike the content and would prefer for it to not exist, they may reasonably conclude that the inclusion of disclaimers exempts it from being legitimately labeled as “defamatory.” Perhaps the issue is as simple as that.

However, I suspect that for many targeted individuals, as well as those who are concerned about their rights, such disclaimers will offer cold comfort. For one thing, not all consumers of

---

<sup>27</sup> Soble points out that some philosophers, such as Helen Longino, have argued for censorship of pornography on the grounds that it is defamatory towards women. Soble rejects this view, arguing that pornography is “nonpropositional” and therefore not defamatory. It is interesting to consider how the specific example of pornographic Deepfake content might complicate Soble’s view about the nonpropositional character of pornography (Soble 1985).

<sup>28</sup> It is worth noting that while Deepfake content often includes both video and audio components, it can consist of video only, as well as audio only. While I do not specifically address audio-only Deepfake content in this paper, my hope is that the arguments presented apply to it in the same way that they apply to other Deepfake content.

media are equally literate, and it remains plausible to think that in some cases, consumers will be deceived into thinking that Deepfake content is authentic even when explicit disclaimers are included. It is reasonable to expect that even when a person is shown a Deepfake video of a prominent individual committing a heinous act, and is informed that the video is inauthentic, the Deepfake content nonetheless collides with their psychology in a way that damages their perception of the individual. It is possible for misinformation to shape individuals' beliefs even in cases wherein they are explicitly informed that a specific piece of content they have been exposed to involves misinformation.

While I am not aware of psychological experiments that involve Deepfake content specifically, the outcomes of experiments involving ordinary video content and print content suggest that misinformation involving Deepfake content can be an especially potent tool for influencing doxastic attitudes.<sup>29</sup> Literature indicates that when people are exposed to misinformation, it is difficult to correct the resulting misbelief. Some strategies for correcting misinformation are more effective than others: issuing a correction before a person is exposed to misinformation appears to be more productive than issuing one after exposure has already taken place. One finding that is particularly concerning is that in some instances, efforts to correct misbelief can actually generate increased misbelief.<sup>30</sup> While we should look forward to thorough psychological research involving Deepfake content that can assist decision-making in this area, it is clear for the time being that the correction of misinformation is far from being a straightforward matter. While the inclusion of disclaimers on Deepfake content may be helpful in some cases, this strategy is by no means a panacea.

---

<sup>29</sup> By "ordinary video content," I simply mean video content that does not involve deployment of Deepfake technology.

<sup>30</sup> Lewandowsky et al. offer a helpful summary of empirical findings related to strategies for correcting misinformation (Lewandowsky et al. 2012).

Another major concern that I wish to raise with regard to the Defamation Argument has to do with the fact that people disagree dramatically about which types of words, images, and ideas are defamatory; the character of defamation is highly indeterminate. This sort of disagreement can be especially pernicious in the context of discussions about Deepfake content and its relationship to individual rights. The concept of defamation ultimately boils down to damage inflicted towards one's reputation.<sup>31</sup> But what constitutes such damage can depend to a large extent on the position that the relevant individual occupies within society, as well as the specific features of this individual. We can probably all agree that a Deepfake video of a person drinking a glass of water lacks the potential to inflict damage towards one's reputation. In addition, we can probably all agree that a Deepfake video of a person abusing someone with a disability does possess the potential to damage their reputation. But there are other cases we might encounter wherein agreement may be difficult to obtain, and determining a common benchmark that can be used as a mechanism for bridging disagreement will be a fraught endeavour.

Consider the example of a Deepfake video that depicts a real person eating beef. To many, this video will clearly lack any sort of defamatory character, as eating beef is considered to be a normal human behaviour that is in no way objectionable. However, members of a specific religious community wherein beef consumption is considered sinful could perceive this video content as being quite scandalous. In cases such as this one, wherein (potential) damage to one's reputation is confined to a specific religious community, how should we go about determining whether an act of defamation has taken place? Does the damage inflicted to an individual's reputation need to be

---

<sup>31</sup> In his discussion of doxing, Douglas outlines a tripartite taxonomy of doxing. He calls the three types of doxing "deanonymizing," "targeting," and "delegitimizing". The last of the three types is defined as doxing that "reveals intimate personal information that damages the credibility of that individual" (Douglas 2016: 199). Since doxing itself is defined by Douglas as "the intentional public release onto the Internet of personal information about an individual by a third party, often with the intent to humiliate, threaten, intimidate, or punish the identified individual," it may be worthwhile to ask whether certain deployments of Deepfake technology, such as the example involving pornographic depictions of celebrities, actually amount to a form of doxing. Douglas specifically cites "involuntary pornography" as an example of delegitimizing doxing, so perhaps it can be argued that pornographic Deepfake content ought to be subsumed under this category.

broad, i.e. impacting the way in which they are generally perceived by society? Or can the damage be more narrow, involving only specific segments of society?<sup>32</sup>

If the latter view is accepted, then we will need to develop an account that can answer yet another question: how narrow is too narrow? While charges of defamation arising from the case I have sketched involving beef consumption may indeed be legitimate, it seems that there are other cases we can imagine (say, wherein a Deepfake video creates a false and embarrassing impression of someone's taste in music) that are just too frivolous to be conceptualized as acts of defamation. Of course, my point here is not to offer an answer regarding which types of Deepfake content are in fact defamatory and which are not. My point is simply that people can and do disagree significantly about what can constitute defamation, and therefore the Defamation Argument carries with it a large philosophical burden. It must be able to handle borderline cases in addition to straightforward cases. This philosophical burden applies to the philosophy of defamation more generally and is not unique to cases involving Deepfake technology. The upshot is that if one is going to endorse the view that Deepfake content can indeed violate the rights of individuals by having a defamatory character, they will not be immune from these concerns regarding the nature of defamation and its inherent indeterminacy.

I anticipate that some will be unconvinced by my criticisms of the Defamation Argument on the grounds that defamation law already exists in the real world, and that it already does a satisfactory job of contending with borderline cases. The sensibility here might be that Deepfake content is just one more venue for defamation on a list of venues that is already long, and that I have exaggerated the significance of borderline cases. There may be merit to this perspective.

---

<sup>32</sup> Peonidis nicely summarizes this point: "...one's conception of defamation is directly dependent on the moral standards that prevail in a specific society, since a person of good reputation is someone who exemplifies certain commonly recognised moral qualities. But even within this particularised context there is room for disagreement as to what constitutes defamation" (Peonidis 1998: 5).

However, I would encourage advocates of this view to bear in mind the considerable degree of heterogeneity that exists with respect to defamation law in various jurisdictions. If one is going to argue that “the law” is already good at handling defamation cases, I will ask which law in which jurisdiction they have in mind. No jurisdiction has a strategy for dealing with defamation cases that is widely agreed to be the most sound. I will leave it to experts on defamation to make the case that existing statutes are already well equipped to handle cases involving Deepfake content, including borderline cases. I remain skeptical that this is the case.

To be clear: the Defamation Argument is far from useless. This argument will be very compelling in the context of a relatively narrow set of cases involving Deepfake technology. There will be deployments of Deepfake technology that are plainly dishonest and malicious, and that are designed to victimize a specific person or group. In such cases, it will be straightforward to see that this content is defamatory. However, it seems that this sort of content will only comprise a relatively small portion of the Deepfake content that is published and circulated in the modern media landscape. While the Defamation Argument will be useful for determining the illegitimacy of Deepfake content that falls into this category, there are many other cases wherein it will struggle to arrive at a clear conclusion about whether a specific act involving Deepfake technology is legitimate. The purpose of this discussion is to identify a reliable framework that can be deployed to assess the legitimacy of Deepfake content in many diverse cases, and not merely in the context of a handful of exceptionally egregious cases. I am searching for a framework that can be applied to content that is malicious, content that is benevolent, and everything in between. This includes Deepfake content that is circulated on obscure online message boards that will only be consumed by a few, as well as Deepfake content published by major Hollywood movie studios that will be consumed by millions. Importantly, due to the fact that this technology is still evolving with respect



to its sophistication and availability, it is important to identify a framework that can accommodate uses of Deepfake technology that we can scarcely imagine, due to the fact that they have not yet taken place. I think that there is an approach to assessing the legitimacy of Deepfake content that is simpler, cleaner, more elegant, and more reliable than the approach offered by the Defamation Argument. This brings us to my discussion of the Property Argument.

#### **(IV) The Property Argument**

The third and final view regarding the relationship between Deepfake technology and individual rights that I wish to consider focuses on property rights. This line of reasoning is based on the foundational premise that individuals own their likenesses as a form of private property.<sup>33</sup> It is for this reason that certain types of Deepfake content can entail a violation of individuals' rights.<sup>34</sup> Let us return to the example of pornographic Deepfake content that exploits the likenesses of celebrities without their consent. Within the framework of the Property Argument, the Deepfake content in question constitutes a violation of the targeted celebrity's rights because it effectively amounts to a form of theft.<sup>35</sup> While individuals (famous or otherwise) can grant permission to Deepfake content producers to use their likenesses for a project, the producers of the content cannot

---

<sup>33</sup> While the issue of ownership of one's body has received attention in philosophy literature, the issue of whether "ownership of one's body" necessarily entails "ownership of one's likeness" does not seem to arise in said literature. Discussions of ownership of one's physical body (and its parts) are offered by Wheeler (Wheeler 1980), Spinello (Spinello 2004) and Björkman & Hansson (Björkman & Hansson 2006). I do not think that the arguments offered by these authors bind them to a specific conclusion with regard to the likeness question. Spinello takes on the issue of property rights in genetic information, and highlights a principle known as "genetic exceptionalism, which asserts that genetic information needs a higher level of protection than other kinds of personal information such as financial data" (Spinello 2004: 29). A point worth considering is whether an individual's likeness ought to be included in the category of "genetic information."

<sup>34</sup> Cwik offers a helpful breakdown of key terms involved in the philosophy of intellectual property: "The term "intellectual property" (or "IP") is a blanket term for a number of legal entitlements, which are collectively referred to as "intellectual property rights" (or "IPRs"). The most familiar of these are patent and copyright; also included in the category are trademark and trade secret protections. Sometimes more exotic legal entitlements (either existing or proposed) such as rights over use of one's public image or likeness, so-called "information rights" (rights to keep certain information, like medical records, private) and the somewhat misleadingly titled "moral rights" of creators of artistic and literary works (to, among other things, attribution as creator of a work) are also grouped with patents, copyrights, trademarks, and trade secrets as IPRs" (Cwik 2016: 471).

<sup>35</sup> Karlen explores the relationship between property rights and aesthetic creations. Since copyright is intertwined with authorship, perhaps one could argue that an individual's likeness ought to be conceptualized as an aesthetic creation that they have authored, and therefore have rightful ownership of. While the relationship between a person's likeness and their own powers of authorship is by no means clear and direct, it might provide a fruitful avenue for inquiry (Karlen 1986).

unilaterally decide to include the likenesses of other real people in their content; doing so lacks legitimacy.<sup>36</sup> An agreement, which may involve financial compensation, must be reached in order for the project to be legitimate.<sup>37</sup>

An advantage of the Property Argument is its conceptual simplicity. I view it as being fairly neat and elegant in comparison to the Privacy Argument and the Defamation Argument. We generally have a good intuitive understanding of the idea that getting something from someone with permission is legitimate, but taking something from them without permission is illegitimate.<sup>38</sup> Inserting personal likenesses into this framework seems like a promising avenue for identifying cases wherein Deepfake technology is being used in a manner that violates the rights of an individual.<sup>39</sup> Some may be skeptical of the notion that individuals can have legal ownership over their own likeness. Surely ownership of a likeness is unlike ownership of other things, such as physical inanimate objects, in some meaningful way. It is fair to suggest that ownership of likenesses is less straightforward than these more mundane cases of ownership. However, my view is that an individual owning their own likeness as a form of private property is not fundamentally different from other forms of intellectual property ownership.

---

<sup>36</sup> Spinello compellingly argues that rights afforded to famous individuals over uses of their likenesses should be extended to all individuals: "... the Courts have traditionally protected famous individuals from having their names or photos used against their will, e.g., to make unauthorized product endorsements. They recognize a property right in these names and photographic images. Why shouldn't a property right also be extended to everyone's *digital persona* (or image) since it too can be used for commercial purposes without authorization?" (Spinello 1998: 734) Of course, with regard to the issue of Deepfake content, I would extend the reasoning further still to include all appropriations of likenesses, even ones that are not being used for commercial purposes.

<sup>37</sup> It is interesting to consider how the Property Argument intersects with the broader debate surrounding data collection and ownership. Since digital data collection is now an immensely lucrative practice, some prominent individuals, including 2020 US Democratic presidential candidate Andrew Yang, have publicly argued that Internet users are entitled to a share of the earnings generated from their data.

<sup>38</sup> McCloskey considers the view that privacy rights actually fall under the umbrella of property rights. Within this paradigm, to have a right to privacy is to own one's privacy as a form of private property. McCloskey rejects the paradigm on the grounds that it leaves open the possibility of individuals consensually buying and selling privacy in a manner that nonetheless constitutes an invasion of privacy. He states: "payment in full of the price demanded is still compatible with gross invasion of privacy, that compensation by way of payment of damages, no matter how great the payment, may be no more real compensation for lack of respect for privacy than is financial compensation for loss of sight or limbs" (McCloskey 1980: 27).

<sup>39</sup> Perhaps one could invoke Locke in order to construct an argument that since individuals do not produce their likenesses through their own labour, it is not legitimate for one to claim their likeness as a form of private property. I am inclined to think that it is plausible to argue that the likenesses of individuals are indeed mixed with their labour since all individuals exert at least some control over their physical appearance. For a discussion of the relationship between intellectual property rights and the deployment of labour see Cwik 2014.

Intellectual property is an integral component of economies around the world, and even though there is room for disagreement about where intellectual property rights begin and end, it is generally accepted that people and groups can have legitimate ownership over abstract (i.e. non-physical) entities. People can and do have property rights over abstract entities such as fictional characters, and in some cases they are willing to pay vast sums of money in order to take on ownership of these entities.<sup>40</sup> If it is reasonable for people to have ownership over the likenesses of fictional characters, then it must be equally reasonable, if not more reasonable, for them to have ownership over their own likeness. After all, fictional characters do not have the ability to become emotionally invested in how their likenesses are used, and there is no reason to worry about whether use of a fictional character's likeness will damage their reputation or livelihood. Meanwhile, actual human beings can easily have their reputation and livelihood damaged if their likeness is used in a manner that is irresponsible or malicious. Such acts can inflict significant suffering upon them. In this sense, ownership of the likenesses of fictional characters is a trivial matter in comparison to ownership of the likeness of a real person. It accordingly makes good sense for us to establish a robust framework that is concerned with how use of people's likenesses is to be governed. My view is that inserting likenesses into a property rights framework is a promising avenue for ensuring that likenesses are not deployed in a manner that is irresponsible or malicious.

The Property Argument does have some features that not all may view as desirable that merit explication. One very significant aspect of this argument is that the specific events being depicted in Deepfake content have no bearing on the force of the argument. I have used the

---

<sup>40</sup> The Walt Disney Company made the decision to purchase Marvel Entertainment for \$4,000,000,000 in 2009, presumably because it reasoned that having the rights to produce and sell content invoking the likenesses of fictional characters such as "Spider-Man," "The Incredible Hulk," and "Iron Man" would be well worth the investment. It is clear that the commodification and exchange of likenesses is something that can and does take place in the real world – sometimes on a scale involving astronomical sums of money.

example of pornographic Deepfake content featuring the likenesses of celebrities throughout this discussion for two reasons: One is that this example has already manifested itself in the real world and generated controversy; it is not merely hypothetical. The other is that this example makes it quite easy to see how and why a person could reasonably find Deepfake content objectionable. If Deepfake technology were only being used for unambiguously innocent or noble purposes, then we might be inclined to view this entire discussion about individual rights as moot. The example involving pornographic depictions of celebrities makes it abundantly clear that some applications of Deepfake technology may be more nefarious than others.

I make this point because the Property Argument is, in essence, completely indifferent towards the types of events being depicted in Deepfake content. Whether the Deepfake content depicts an individual committing a heinous crime, or whether it merely depicts them performing some mundane act, it cannot evade the scope of the Property Argument. Within this framework, appropriating the likeness of a person is illegitimate if consent from the person is not provided, and this illegitimacy stands regardless of whether the events depicted in the Deepfake content are good or bad or neutral. Some may view this as a defect of the Property Argument, but I am inclined to view it as a desirable feature. This simplicity enables us to avoid burdensome philosophical questions regarding personal impact, and whether this or that depiction is sufficiently damaging in order to constitute a rights violation. This relative parsimony is a virtue of the Property Argument.

In the preceding section, I noted that there will be deployments of Deepfake technology that clearly constitute defamation. One may plausibly wonder how I can admit this point, yet still defend the conclusion that the Property Argument offers a superior approach to the regulation of Deepfake content than the Defamation Argument. The answer is simply that the Property Argument has a much greater scope than the Defamation Argument. While it is true that the

Defamation Argument can be just as powerful as the Property Argument in identifying illegitimate deployments of Deepfake technology in some cases, it is much less powerful when we take a more general view of Deepfake content. Let us suppose that we want to inquire about the legitimacy/illegitimacy of all of the Deepfake content that exists. If we adopt the Defamation Argument, we will need to make distinctions between various types of content based on how threatening they are to the reputation and livelihood of the person whose likeness is used in the content. A plausible classificatory scheme that seeks to organize all of the Deepfake content that exists may invoke the following five categories: 1.) Clearly Defamatory 2.) Probably Defamatory 3.) Possibly Defamatory 4.) Probably Not Defamatory 5.) Clearly Not Defamatory. While the Defamation Argument can easily produce conclusions regarding prohibition of content that falls into Category 1, for every other category its power to regulate content is either limited or nonexistent. Even though there may be very good reasons to be concerned about content that falls into categories 2 through 5, the Defamation Argument will be ill-equipped to produce actionable prescriptions regarding the prohibition of this content.

In order to illustrate the limitations of the Defamation Argument in comparison to the Property Argument, let us once again consider the example of pornographic Deepfake content that exploits the likenesses of celebrities. More specifically: let us imagine that a Deepfake video has been published online that depicts a celebrity engaging in consensual sexual activities with their real-life spouse, and that these activities include nothing out of the ordinary. Of the five categories listed above, which category does this content belong in? I find it difficult to produce a confident answer here. I can imagine plausible arguments being offered about why this content belongs in each of the categories I have enumerated. There is nothing controversial about a celebrity engaging in consensual sexual relations with their spouse, and therefore one may argue that this content

belongs in Category 5. Meanwhile, the targeted celebrity may plausibly argue that the content belongs in Category 1 since it depicts them in a manner that they would never be comfortable sharing with the public, and that they believe has compromised their public perception and image. If we adopt the Defamation Argument and use it to establish a framework for the regulation of Deepfake content, then these types of debates can continue indefinitely.

Alternatively, the Property Argument can entirely sidestep these sorts of complicated and taxing debates. In order to invoke the Property Argument, we need to ask exactly one question: did the Deepfake content in question use a person's likeness without their consent? If the answer is yes, then we have grounds to disallow this content. The five categories identified above are all fair game for this argument to an equal extent. It is worthwhile to clarify the conclusion that these observations lead toward. Even though the Defamation Argument can provide an equally strong case for placing limits upon Deepfake content in a relatively narrow set of cases, in general, the Property Argument offers a clearer and more robust mechanism for arriving at conclusions about the legitimacy and illegitimacy of Deepfake content. This paper is interested in identifying an approach to Deepfake technology and individual rights that is useful in a broad variety of cases, and not just in some exceptionally clear-cut cases, and that is why I defend the conclusion that the Property Argument is superior to the Defamation Argument, as well as the Privacy Argument.

A reasonable question to ask at this point is whether the Property Argument forbids practices that are ubiquitous and widely viewed as legitimate. When a person does something that is newsworthy, it is usually the case that news outlets proceed to publish the likeness of this person without ever requesting permission. This seems like an unproblematic practice that would fall afoul of the Property Argument, and this may be used as a rationale for rejecting the argument. I

think that the Property Argument can be defended from this sort of challenge, and I will identify a strategy for achieving this.

The Property Argument can be protected from this challenge by introducing a public interest criterion that functions as a check on the argument. This public interest criterion will hold that exceptions can be made within the context of the Property Argument if and when it is determined that publishing the likeness of a person who has not granted consent is vital to the public interest. An uncontroversial example will involve a known assailant who has not yet been apprehended by authorities. It is intuitive and sound to suggest that it is legitimate for media outlets to circulate the likeness of this person without their consent, due to the fact that it is in the public's interest to be familiar with the appearance of this person for the sake of safety. The reasoning here will maintain that while this assailant does indeed possess a property right with respect to their likeness, it is acceptable for other parties to override this right when doing so fulfills a vital public interest. Publishing the appearance of this person is accordingly a limitation on the relevant property right rather than a violation of the right.<sup>41</sup>

This is a view that should be wholly inoffensive to most, and this includes committed liberals. Even the staunchest defenders of individual rights generally admit that rights are not absolute, and that certain limits can be placed on these rights. The right to free expression does not entitle one to falsely report a bomb threat at a crowded airport. The right to privacy does not entitle one to enjoy solitude in their apartment unit when authorities have reason to think that they are conducting a methamphetamine operation that could cause fires and explosions that will jeopardize

---

<sup>41</sup> Note that while ownership over likenesses is already commonplace in the world of intellectual property, this type of ownership comes with certain limitations. In US copyright law, the "fair use" doctrine empowers individuals and groups who do not have ownership over a piece of intellectual property to make use of it without first obtaining permission from the relevant right holder. While there is room for disagreement over what exactly qualifies as fair use, the doctrine is successfully invoked on a routine basis. An implication is that even after paying billions of dollars for the right to use the likenesses of certain fictional characters, powerful companies such as the Walt Disney Company must accept that others will continue to use these likenesses so long as they adhere to rules regarding fair use.

the lives of others living in the same apartment complex. Likewise, a property right over one's own likeness does not entitle one to inflict harms on others without having this property right limited so that these harms can be mitigated. To a certain extent, a liberal framework presupposes that individual rights can be limited when they come into conflict with the rights of others, and it should not be surprising or bothersome that property rights over likenesses involve certain qualifications and limitations.

It is also worth noting that no feature of the Property Argument necessarily precludes individuals from opting out of private ownership over their likeness. We can easily imagine a service being provided that enables individuals to declare publicly whether or not they consent to their likeness being used by others without direct, explicit permission. This could streamline the permission acquisition process and reduce the frequency with which parties must communicate directly with individuals in order to obtain consent to employ their likeness. Like any other property right, a property right over one's likeness can be waived or forfeited.

This discussion of the strengths and weaknesses of the Property Argument is far from exhaustive. Given the level of sophistication of Deepfake technology, and the variety of ways in which it may be used, there is no shortage of possible cases involving the technology that are philosophically interesting from the standpoint of individual rights.<sup>42</sup> Perhaps there are cases that I have not yet considered that will present serious challenges for this argument.<sup>43</sup> The overarching question that animates this discussion is whether there are compelling arguments in favour of

---

<sup>42</sup> As was noted earlier, the case of *Aubry v Éditions Vice-Versa Inc* in Quebec highlights the distinction between producing and publishing a photograph of an individual without their consent. This distinction between producing and publishing seems like it also is relevant within the context of the Property Argument. Is it legitimate to appropriate another's likeness without permission for the sake of producing Deepfake content that is intended solely for individual personal use? It seems that the Property Argument could be formulated in order to answer this question in the negative or the affirmative. I will not attempt to answer this question here.

<sup>43</sup> One issue that will need to be addressed involves different individuals who are physically identical, or near-identical. We can imagine a case involving identical twins, wherein one consents to their likeness being used in Deepfake content while the other abstains from granting consent. Has the former twin here violated the rights of the latter? I am inclined to think that this would not amount to a formal, substantive violation of the twin's rights, and that we will need to bite the proverbial bullet and accept that the property argument cannot shield all individuals from deployments of Deepfake content that are offensive and/or morally dubious.



constraining the deployment of Deepfake technology on the grounds that its use violates the rights of individuals. Currently, my view is that the Property Argument lays out the most robust and reliable strategy for legitimately constraining Deepfake content in the interest of protecting individual rights.

It is worth pointing out that property rights are an integral component of the classical liberal tradition,<sup>44</sup> as is the right to free expression. This is significant because concerns about illiberalism can be advanced by individuals on both sides of the broader debate about Deepfake technology and whether it can be legitimately constrained. Individuals who think that the production and distribution of Deepfake content should be completely unrestricted can argue that those on the opposing side of the debate are insufficiently committed to the right to free expression. Meanwhile, those who endorse the Property Argument and accordingly think that placing constraints on Deepfake content can be legitimate, can argue that their opponents are insufficiently committed to property rights. It is not obvious which orientation is more faithfully aligned with the liberal tradition. Deepfake technology thus presents an opportunity for an exceptionally interesting case study in applications of liberal philosophy and theory more generally. My hope is that a richer literature on the subject is on the horizon.

### **(V) Marketplace Solutions**

I wish to supplement my discussion of Deepfake technology and individual rights by briefly acknowledging that there are strategies for addressing problematic uses of Deepfake technology that do not necessarily implicate rights (or rights theory). Let us consider once again

---

<sup>44</sup> Interestingly, in this regard, property rights stand in especially stark contrast to privacy rights. McCloskey notes that major classical liberal thinkers such as Locke and Mill do not take on the issue of privacy rights in their writings. He argues that contemporary notions about the right to privacy have their roots in the American legal profession rather than the work of philosophers (McCloskey 1980).

the example of pornographic Deepfake content that exploits the likenesses of celebrities without their consent. Issues of rights aside, it may simply turn out to be the case that consumers arrive at a consensus that this type of content is morally unacceptable, and that any person or company involved in the production or distribution of such content will pay a high cost as a result. It may be public backlash, rather than charges of rights violations, that proves most efficacious with regard to regulating the use of Deepfake technology. In 2018, the website Reddit shut down an online community dedicated to pornographic Deepfake content that had existed on its platform, suggesting that private companies may be able and willing to regulate this content in a manner that is satisfactory to those who are concerned about it.

Indeed, a number of the world's most popular online platforms have stated that pornographic Deepfake content that appropriates celebrity likenesses violates their rules prohibiting non-consensual (or involuntary) pornography, and accordingly will work to remove this type of content if it is detected on their platform (Cole 2018). This development seems like an important step in the right direction. However, this development should not be taken as an indication that such content is on its way to being purged from the marketplace, as it remains entirely possible that a niche market will emerge that can remain profitable without cooperation from these large digital media platforms. Perhaps the market for Deepfake pornography featuring celebrity likenesses will run parallel to the mainstream pornography market and the two will remain separate. It remains possible that producers will be able to profit significantly from a small number of consumers who are willing to pay high prices for pornographic content including the likeness of their preferred celebrity. The upshot is that even if every mainstream digital media platform agrees that this type of Deepfake content is unacceptable, this does not guarantee that this content will not be readily available to those who want to access and purchase it.

It is of course also worth noting that one of the most prominent fears regarding Deepfake technology concerns its application in political propaganda and disinformation campaigns.<sup>45</sup> This topic is far too rich to treat in detail here. However, it is worth taking a moment to consider whether the marketplace will have an answer for this very legitimate concern. It seems to me that in most cases, a major media outlet that commits the act of distributing such content to the public in a manner that misinforms<sup>46</sup> it will accordingly suffer a very severe loss of credibility. This loss of credibility may prove damaging or even fatal to their fiscal bottom line. Credibility is an important form of currency in the media landscape. While a great deal could be said about journalistic failures that have been committed by large media outlets that have continued to operate, my general view is that there are journalistic failures so egregious that they can spell the demise of an institution. I expect that perpetuating falsehoods through Deepfake content would constitute one such journalistic failure. Perhaps media outlets will be so protective of their credibility in the wake of widespread dissemination of Deepfake content that it will actually help raise the proverbial bar in the world of journalism, and lead to greater vigilance in separating fact from fiction.

While there is room for some optimism in this debate, we must also recognize the possibility that the news media marketplace will operate in a manner analogous to the one outlined above involving a niche market for pornographic content involving celebrity likenesses. While news outlets that conform to journalistic norms involving fact-checking and error correction will tread carefully with respect to Deepfake content in order to protect their credibility, there will

---

<sup>45</sup> A potentially surprising aspect of applications of Deepfake technology that are politically motivated is that these applications can be argued to have an especially strong claim to legitimacy as forms of expression, in comparison to other types of content employing Deepfake technology. A precedent for this view can be found in the United States. The US Supreme Court has undertaken the project of distinguishing between various categories of speech, and has concluded that some categories (such as political speech) enjoy more robust First Amendment protections than other categories (such as commercial speech). Accordingly, if some Deepfake content is determined to fall under the umbrella of “political speech,” then this content will actually be more difficult to constrain (at least in the US context) than other types, such as pornographic content. For a more detailed discussion of speech categories and their First Amendment protections, see Sumner 2004.

<sup>46</sup> Misinformation through social media has received philosophical treatment from Millar (Millar 2019), Obadã (Obadã 2019) and Pepp, Michaelson, and Sterken (Pepp, Michaelson, and Sterken 2019).

likely be other outlets that profit from the dissemination of misinformation and refuse to concede that such content is inauthentic, thereby catering to an audience that is willing (even eager) to consume misinformation.<sup>47</sup> Today we see segments of the Internet wherein users are effectively partitioned off from credible sources of information. These users participate in echo chambers wherein they descend increasingly deeply into narratives that are divorced from reality. While the marketplace mechanisms described above may be useful for constraining misinformation involving Deepfake content in contexts wherein users are not fully siloed off from credible sources that adhere to journalistic norms, they will likely struggle to reign in misinformation in contexts where media outlets and their audiences are already deeply steeped in falsehood and are committed to narratives that deem conventional information sources untrustworthy.<sup>48</sup> These actors may simply refuse to recognize misinformation as such, no matter what fact-checking bodies have to say about the matter.<sup>49</sup> In such contexts, the best outcome one can hope for with respect to marketplace solutions is that consumers of misinformation will develop some willingness to become exposed to alternative outlets and arguments that are more grounded in reality, and ultimately decide to distance themselves from sources of misinformation. It is easy to see why many would be pessimistic about the prospect of this taking place on any sort of large scale in the near future.

I will reiterate that these comments about marketplace solutions are intended to be supplementary to my discussion about individual rights. I do not wish to give the impression that market-oriented solutions are clearly superior to approaches invoking rights. Indeed, it is important

---

<sup>47</sup> It is important to bear in mind that there is a distinction between two different categories of misinformation outlets: those that disseminate misinformation because they sincerely believe they are telling the truth, and those that disseminate misinformation, and know that they are disseminating misinformation, but do so anyway because it serves their financial interests. My discussion here is inclusive of both categories.

<sup>48</sup> For a more detailed discussion of how misinformation/disinformation are corrected in the modern media ecosystem, as well as how they are permitted to thrive, see Benkler, Faris, and Roberts 2018.

<sup>49</sup> I would like to thank an anonymous reviewer for raising this point about the volume of public demand for misinformation.

that philosophers and political theorists work to establish a rights framework that can accommodate the rapidly proliferating phenomenon of Deepfake content, and that this work inform future decisions by legislators. However, I am comfortable with espousing a two-pronged approach to addressing issues involving Deepfake content that highlights the importance of both rights theory as well as marketplace solutions, rather than choosing one over the other.

It is important to note that in the absence of rights theory, marketplace solutions will likely involve a significant degree of arbitrariness and selectivity. Indeed, this sort of dynamic has already manifested itself in the context of social media platforms and their decisions regarding removal of certain types of content. While many agree that it is legitimate for private social media companies to filter out certain types of content (such as misinformative and hateful content), decisions to remove content are often met with considerable skepticism, confusion, and even anger as users point out inconsistencies in how these decisions are implemented. Marketplace solutions may thus be appropriately conceptualized as supplementary strategies for filtering out problematic content from online platforms that can complement work being done in the domain of rights theory. My hope is that this work will eventually shape legislation in beneficial ways.

### **(VI) Concluding Remarks**

No doubt some readers will notice an issue that is conspicuously absent from my discussion: the distinction between ordinary individuals and public figures. I understand that in certain jurisdictions there are legal distinctions that separate public figures such as celebrities from the rest of the population. Much could be said about the concept of the public figure and its philosophical implications. I will here limit myself to expressing my skepticism as to whether this concept is tenable in light of recent developments in technology and culture. In an age wherein

many millions of individuals are routinely publicizing photos and videos of themselves (in addition to other details about their lives) via social media in hopes of garnering sizable followings, it becomes much more difficult to determine who exactly ought to be categorized as a “public figure.” While at one time the question of who is and is not a public figure seemed relatively clear-cut, I do not believe that this is the case any longer.

Within the context of contemporary culture, while many individuals may not be celebrities per se, they operate in a manner similar to celebrities in that they use digital media as a vehicle for establishing and promoting a personal brand. At this moment, the number of social media users who identify themselves as “influencers” and “brand ambassadors” continues to proliferate. While this type of activity can generate income for social media users, this income can be supplementary; not everyone who profits financially from social media does so on a full-time basis. Even individuals with perfectly ordinary professional and personal lives are sharing content online in hopes of it getting traction and attention from others whom they have never encountered face to face. While social media platforms generally give users the option to keep their profiles private, many choose to make their profiles public, meaning that anyone with an Internet connection can access the content they choose to share. Given that this dynamic is now ubiquitous, and only increasing in terms of breadth and depth, I have serious doubts as to whether any authority can determine in a sound fashion which segment of a population consists of “public figures” and which segment does not. In the modern digital age, the traditional concept of “the public figure” is more problematic than is often recognized, and that is why I have sought to avoid it completely throughout the preceding discussion.

Finally, I must clarify that my view regarding the relative superiority of the Property Argument over the Defamation Argument and the Privacy Argument is not intended to preclude

the possibility of espousing a hybrid approach wherein the strongest elements of each argument are synthesized in some way. I have not undertaken this project here simply because it would require a great deal more space than I have available. This discussion is only the first step towards formulating practical solutions in a world that is just beginning to adjust to the new reality of Deepfake technology. Indeed, my discussion has been animated by a sense that this matter is somewhat pressing, and that is why I have sought to propose strategies for grappling with the issue that can bypass peripheral philosophical debates and establish a philosophical rights framework that is capable of efficiently determining the legitimacy and illegitimacy of various applications of Deepfake technology. I have expressed partiality towards the Property Argument because it offers a rights framework that is relatively simple and parsimonious. My hope is that this framework will prove valuable as stakeholders move to identify and address problematic deployments of Deepfake technology.

### **Acknowledgements**

I am grateful to the faculty and graduate students in the McMaster University Philosophy Department who participated in the 2019-2020 PhD Research Seminar for providing feedback about an earlier version of this paper. Their interest and encouragement helped motivate me to continue to refine my views about the intersection between Deepfake technology and individual rights.



## References

- Anderson, Scott A. 2008. "Privacy Without the Right to Privacy," *Monist: An International Quarterly Journal of General Philosophical Inquiry* 91 (1): 81-107.
- Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.
- Björkman, Barbro and Sven Ove Hansson. 2006. "Bodily rights and property rights," *Journal of Medical Ethics: The Journal of the Institute of Medical Ethics* 32 (4): 209-214.
- Cole, Samantha. February 6, 2018. "Twitter Is the Latest Platform to Ban AI-Generated Porn," *Motherboard (Tech by Vice)*. [https://www.vice.com/en\\_us/article/ywqgab/twitter-bans-deepfakes](https://www.vice.com/en_us/article/ywqgab/twitter-bans-deepfakes)
- Cole, Samantha. Video published February 7, 2020. "What are Deepfakes? The Real Implications of Fake Porn". <https://www.youtube.com/watch?v=FqzE6NOTM0g&feature=youtu.be>
- Cwik, Bryan. 2014. "Labor as the basis for intellectual property rights," *Ethical Theory and Moral Practice: An International Forum* 17 (4): 681-695.
- Cwik, Bryan. 2016. "Property rights in non-rival goods," *Journal of Political Philosophy* 24 (4): 470-486.
- Douglas, David M. 2016. "Doxing: A conceptual analysis," *Ethics and Information Technology* 18 (3): 199-210.
- Feinberg, Joel. 1984. *The Moral Limits of the Criminal Law: Harm to Others*. New York: Oxford University Press.
- Feinberg, Joel. 1985. *The Moral Limits of the Criminal Law: Offense to Others*. New York: Oxford University Press.

- Janeček, Václav. 2018. "Ownership of personal data in the Internet of Things," *Computer Law and Security Review* 34 (5): 1039-52.
- Karlen, Peter H. 1986. "Worldmaking: Property Rights in Aesthetic Creations," *Journal of Aesthetics and Art Criticism* 45: 183-192.
- Lewandowsky et al. 2012. "Misinformation and Its Correction: Continued Influence and Successful Debiasing," *Psychological Science in the Public Interest* 13 (3): 106-131.
- Marx, Gary T. 2016 *Windows into the Soul: Surveillance and Society in an Age of High Technology*. Chicago: University of Chicago Press.
- McCloskey, H.J. 1980. "Privacy and the Right to Privacy," *Philosophy: The Journal of the Royal Institute of Philosophy* 55 (211): 17-38.
- Metz, Rachel. September 5, 2019. "Facebook Is Making Its Own Deepfake Videos to Help Fight Them." *CNN*, Cable News Network, [www.cnn.com/2019/09/05/tech/facebook-deepfake-detection-challenge/index.html](http://www.cnn.com/2019/09/05/tech/facebook-deepfake-detection-challenge/index.html).
- Mill, John Stuart. 2015. *On Liberty, Utilitarianism, and Other Essays*. New York: Oxford University Press.
- Millar, Boyd. 2019. "The Information Environment and Blameworthy Beliefs," *Social Epistemology* 33 (6): 525-537.
- Nakashima, Ryan. August 13, 2018. "Google Tracking Your Movements, Regardless of Privacy Settings." *CTV News*, Canadian Television. [www.ctvnews.ca/sci-tech/google-tracking-your-movements-regardless-of-privacy-settings-1.4050286](http://www.ctvnews.ca/sci-tech/google-tracking-your-movements-regardless-of-privacy-settings-1.4050286).
- Obadă, Rareș. 2019. "Sharing Fake News about Brands on Social Media: a New Conceptual Model Based on Flow Theory," *Argumentum. Journal of the Seminar of Discursive Logic, Argumentation Theory and Rhetoric* 17 (2): 144-166.

- Peonidis, Filimon. 1998. "Freedom of expression, autonomy, and defamation," *Law and Philosophy: An International Journal for Jurisprudence and Legal Philosophy* 17 (1): 1-17.
- Pepp, Jessica, Eliot Michaelson & Rachel Katharine Sterken. 2019. "What's New About Fake News?," *Journal of Ethics and Social Philosophy* 16 (2): 67-94.
- Rosebury, Brian. 2011. "The Political Logic of Victim Impact Statements," *Criminal Justice Ethics* 30 (1): 39-67.
- Schoeman, Ferdinand David. 1992. *Privacy and Social Freedom, (Cambridge Studies in Philosophy and Public Policy)*. Cambridge: Cambridge University Press.
- Soble, Alan. 1985. "Pornography: Defamation and the Endorsement of Defamation," *Social Theory and Practice: An International and Interdisciplinary Journal of Social Philosophy* 11 (1): 61-88.
- Spinello, Richard A. 1998. "Privacy Rights in the Information Economy," *Business Ethics Quarterly* 8 (4): 723-742.
- Spinello, Richard A. 2004. "Property rights in genetic information," *Ethics and Information Technology* 6 (1): 29-42.
- Sumner, L.W. 2004. *The Hateful and the Obscene: Studies in the Limits of Free Expression*. University of Toronto Press.
- Thomson, Judith Jarvis. 1975. "The Right to Privacy," *Philosophy & Public Affairs* 4 (4): 295–314.
- Véliz, Carissa. 2020. *Data, Privacy, and the Individual*. Center for the Governance of Change (2020) Available on PhilArchive: <https://philarchive.org/archive/VLIPM>

Vold, Karina & Jessica Whittlestone. 2019. "Privacy, Autonomy, and Personalised Targeting: Rethinking How Personal Data is Used," Appears in Carissa Véliz (ed.), *Data, Privacy, and the Individual*. Center for the Governance of Change.

Wheeler, Samuel C. 1980. "Natural Property Rights as Body Rights," *Noûs* 14 (2): 171-194.

**Information about Author**

**Author name:** Francesco Stellin Sturino

**Affiliation:** Department of Philosophy, McMaster University

**Email address:** sturinof@mcmaster.ca