

University of Miami Law School

University of Miami School of Law Institutional Repository

Articles

Faculty and Deans

2019

Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions

Mary Anne Franks

Ari Ezra Waldman

Follow this and additional works at: https://repository.law.miami.edu/fac_articles



Part of the [First Amendment Commons](#), and the [Law and Society Commons](#)

SEX, LIES, AND VIDEOTAPE: DEEP FAKES AND FREE SPEECH DELUSIONS

MARY ANNE FRANKS* & ARI EZRA WALDMAN**

“If you are neutral in situations of injustice, you have chosen the side of the oppressor.”—Desmond Tutu¹

The longstanding position of civil libertarians that harmful speech should generally be tolerated instead of regulated² is based on three interrelated claims about free speech. One is that an unfettered “marketplace of ideas” ultimately leads to the discovery of truth.³ The second, closely related to the first, is that harmful speech is always best addressed through counter-speech rather than regulation.⁴ The third is that even well-intentioned and modest regulations of speech will ultimately be used to silence minority or dissident voices.⁵ Whatever merit these claims may have had in the past, they cannot be sustained in the digital age. Unbridled, unlimited free speech rights, especially in an era of technologically mediated expression, have led to the disintegration of truth, the reign of unanswerable speech, and the silencing and self-censorship of women, queer people, persons of color, and other racial and ethnic minorities.⁶

© 2019 Mary Anne Franks & Ari Ezra Waldman.

* Professor of Law, University of Miami School of Law; President and Legislative and Tech Policy Director, Cyber Civil Rights Initiative; D.Phil., Oxford University; J.D. Harvard Law School.

** Professor of Law and Founding Director, Innovation Center for Law and Technology, New York Law School; Fellow, Information Society Project, Yale Law School. Ph.D., Columbia University; J.D., Harvard Law School. The authors would like to thank Danielle Keats Citron for her leadership and organizing a symposium on deep fakes. Thanks also to the symposium participants, including Bobby Chesney, Stacey Dogan, Woodrow Hartzog, Quinta Jurecic, Thomas Kadri, Kate Klonick, Jessica Silbey, Olivier Sylvain, and Benjamin Wittes. The authors share credit and responsibility for this essay equally. Special thanks to the student editors of the *Maryland Law Review*.

1. Gary Younge, *The Secrets of a Peacemaker*, GUARDIAN (May 22, 2009), <https://www.theguardian.com/books/2009/may/23/interview-desmond-tutu>.

2. See generally ANTHONY LEWIS, *FREEDOM FOR THE THOUGHT WE HATE* (2007); NADINE STROSSEN, *HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP* (2018).

3. See generally Thomas W. Joo, *The Worst Test of Truth: The “Marketplace of Ideas” as Faulty Metaphor*, 89 TUL. L. REV. 383 (2014) (critiquing this claim).

4. See STROSSEN, *supra* note 2, at 130.

5. *Id.* at 81 (arguing that equality for marginalized groups depends on free speech); Mary Ellen Gale & Nadine Strossen, *The Real ACLU*, 2 YALE J.L. & FEMINISM 161, 180 (1989) (noting that dissident voices, on the left and right, are the ones targeted by censorship).

6. See, e.g., DANIELLE KEATS CITRON, *HATE CRIMES IN CYBERSPACE* (2014); Ari Ezra Waldman, *Law, Privacy, and Online Dating: “Revenge Porn” in Gay Online Communities*, 44 L.

These three delusions of traditional free speech discourse are thrown into sharp relief by the phenomenon of “deep fakes.” Deep fakes, a portmanteau of “deep-learning” and “fake,” are audio or visual material digitally manipulated to make it appear that a person is saying or doing something that they have not really said or done.⁷ First, deep fakes erode the capacity of the public to discern truth from falsity. Second, they inflict dignitary harm that cannot be challenged or corrected by responsive expression. Third, it is the *failure* to regulate deep-fake pornography—not the efforts to do so—that most seriously undermines the free speech rights of vulnerable groups. To do nothing about harmful speech in the digital age is far from liberal nonintervention;⁸ rather, it is a normative choice that perpetuates the power of entrenched majorities against vulnerable minorities.

Deep-fake technology leverages machine-learning algorithms to create highly realistic impersonations from images or audio recordings of actual people.⁹ The technology has been used most notoriously to create hardcore pornographic videos that transplant the faces of famous women onto the bodies of others.¹⁰ Deep-fake pornography, therefore, is closely related to what is often colloquially referred to as “revenge porn,” or nonconsensual pornography,¹¹ the disclosure of private, intimate images or videos of another person, without that person’s consent.¹² Like other forms of nonconsensual pornography, digitally manipulated pornography turns individuals into objects of sexual entertainment against their will, causing intense distress, humiliation, and reputational injury.¹³ Research to date indicates that the targets of

& SOC. INQUIRY (forthcoming 2019); Mary Anne Franks, *Sexual Harassment 2.0*, 71 MD. L. REV. 655 (2009).

7. Robert Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALF. L. REV. (forthcoming 2019) (manuscript at 4), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954.

8. For an in-depth discussion of liberal nonintervention, please see JOHN RAWLS, *A THEORY OF JUSTICE* (1971) and John Rawls, *Justice as Fairness: Political Not Metaphysical*, 14 PHIL. & PUB. AFF. 223 (1985).

9. See Chesney & Citron, *supra* note 7 (manuscript at 4).

10. Drew Harwell, *Fake-Porn Videos Are Being Weaponized to Harass and Humiliate Women: “Everybody Is a Potential Target,”* WASH. POST (Dec. 30, 2018), https://www.washingtonpost.com/technology/2018/12/30/fake-porn-videos-are-being-weaponized-harass-humiliate-women-everybody-is-potential-target/?utm_term=.eaf14bf0e0bf.

11. See Mary Anne Franks, *“Revenge Porn” Reform: A View from the Front Lines*, 69 FLA. L. REV. 1251 (2018) [hereinafter Franks, *“Revenge Porn” Reform*]. The essential evil of revenge porn is not the motive animating the behavior, but the invasion of privacy and the transformation of victims into objects without their consent. See Mary Anne Franks, *How to Defeat ‘Revenge Porn’: First, Recognize It’s About Privacy, Not Revenge*, HUFFINGTON POST (June 22, 2015), http://www.huffingtonpost.com/mary-anne-franks/how-to-defeat-revenge-porn_b_7624900.html.

12. Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345, 346 (2014).

13. See CITRON, *supra* note 6, at 6–10; Franks, *“Revenge Porn” Reform*, *supra* note 11, at 1259–60.

nonconsensual pornography tend to be disproportionately female¹⁴ as well as disproportionately queer.¹⁵

Deep-fake manipulation is, at its most fundamental, a form of deliberately deceptive speech. As such, its unchecked proliferation is completely at odds with a society that values the pursuit of truth. Civil libertarians argue that revenge porn, fake news, and other forms of bad speech are part of the price we pay for a free society.¹⁶ It is part of the “marketplace of ideas,”¹⁷ arguably one of the most powerful governing analogies in First Amendment law.¹⁸ The marketplace metaphor suggests that a functioning democracy should allow its citizens to determine the best idea among many by letting ideas fight it out in the field: “[T]he best test of truth is the power of the thought to get itself accepted in the competition of the market,” Holmes wrote.¹⁹ In this view, good ideas, like the best products, will win out and bad ideas, like inferior, faulty, or poorly made products, will be tossed aside. Getting the state involved in any capacity in the fight against fake news, the argument goes, runs afoul of this most basic free speech principle of nonintervention and tolerance.

But deliberately deceptive speech undermines, not enhances, the pursuit of truth. What is more, the argument that truth will ultimately emerge from

14. According to a 2017 study conducted by the Cyber Civil Rights Initiative, women are 1.7 times more likely to be victims than men. ASIA A. EATON ET AL., CYBER CIVIL RIGHTS INITIATIVE, 2017 NATIONWIDE ONLINE STUDY OF NONCONSENSUAL PORN VICTIMIZATION AND PERPETRATION, A SUMMARY REPORT 12 (2017), <https://www.cybercivilrights.org/wp-content/uploads/2017/06/CCRI-2017-Research-Report.pdf>; see Anastasia Powell et al., *The Picture of Who Is Affected by ‘Revenge Porn’ Is More Complex Than We First Thought*, CONVERSATION (May 7, 2017), <https://theconversation.com/the-picture-of-who-is-affected-by-revenge-porn-is-more-complex-than-we-first-thought-77155> (noting that “there are many more sites and platforms dedicated to sharing women’s nude or sexual images without their consent than men’s”); AMANDA LENHART ET AL., DATA & SOC’Y RESEARCH INST., NONCONSENSUAL IMAGE SHARING: ONE IN 25 AMERICANS HAS BEEN A VICTIM OF “REVENGE PORN” 5 (2016), https://datasociety.net/pubs/oh/Nonconsensual_Image_Sharing_2016.pdf (finding that one in ten women under the age of thirty had been threatened with disclosure of intimate images); Abby Whitmarsh, *Analysis of 28 Days of Data Scraped from a Revenge Pornography Website*, WORDPRESS.COM (Apr. 13, 2015), <https://everlastingstudent.wordpress.com/2015/04/13/analysis-of-28-days-of-data-scraped-from-a-revenge-pornography-website/> (finding that of 396 posts to a revenge porn website, 378 depicted women versus 18 men).

15. According to Data & Society’s 2016 Report, individuals who identified as lesbian, gay, or bisexual were more than seven times as likely to be threatened with nonconsensual pornography than heterosexual individuals. LENHART ET AL., *supra* note 14, at 5.

16. See Ari Ezra Waldman, *The Marketplace of Fake News*, 20 U. PA. J. CONST. L. 845, 847 (2018).

17. 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

18. See, e.g., Joseph Blocher, *Institutions in the Marketplace of Ideas*, 57 DUKE L.J. 821, 829–32 (2008) (examining the Supreme Court’s remedial decision to accord more speech when falsehoods or bad ideas persist).

19. *Abrams*, 250 U.S. at 630 (Holmes, J., dissenting).

the market has been debunked²⁰: Studies show that truthful speech commonly gets drowned out by fake, misleading, and bad content.²¹ Attempts to correct untruths often backfire, due to a phenomenon known as “the illusory truth effect.”²² Repeated exposure to false information, even when presented for the purposes of correction, increases the likelihood that the false information will be remembered as true.²³ Studies show that people are more likely to accept false headlines, for example, when they encounter them more than once.²⁴

Not only is it difficult to correct false ideas, but some false ideas are effectively unanswerable. There is no real way to speak back to a fraudulent representation that is virtually indistinguishable from a real depiction of an individual engaged in graphic intimate activity. Like the unauthorized publication of a person’s actual nude image, the dissemination of a home address, or the disclosure of one’s sexual orientation, deep fakes are not ideas that can simply be countered with different and better ideas.

Finally, deep fakes undermine free speech itself, at least of its targets. Deep fakes weaponize targets’ speech against themselves, harvesting their photos, videos, and audio recordings to create increasingly realistic, fraudulent representations.²⁵ Like other forms of nonconsensual pornography, deep fakes erode the trust that is necessary for social relationships²⁶ and political

20. See, e.g., Lynn Hasher et al., *Frequency and the Conference of Referential Validity*, 16 J. VERBAL LEARNING & VERBAL BEHAV. 107, 111–12 (1977) (explaining that people are heavily influenced by the first version of a story they hear); see also Norbert Schwarz et al., *Metacognitive Experiences and the Intricacies of Setting People Straight: Implications for Debiasing and Public Information Campaigns*, 39 ADVANCES EXPERIMENTAL SOC. PSYCHOL. 127, 152 (2007) (reaffirming through research that memory does not serve to distinguish between true and false details).

21. See Lisa K. Fazio et al., *Knowledge Does Not Protect Against Illusory Truth*, 144 J. EXPERIMENTAL PSYCHOL.: GEN. 993, 993–1002 (2015) (finding that “[c]ontrary to prior suppositions, illusory truth effects occurred even when participants knew better”); Joo, *supra* note 3, at 408 (“If consumers do not value truth very much (relatively speaking), perfect competition will efficiently ensure that they don’t get very much truth as compared with other goods”); Joe Concha, *Pew Study Finds Americans Can’t Tell Fact from Opinion*, THE HILL (June 18, 2018), <https://thehill.com/homenews/media/392870-pew-study-finds-americans-cant-tell-fact-from-opinion> (one example: “Nine-in-10 Democrats correctly identified the statement ‘President Barack Obama was born in the United States’ as factual, while only 63 percent of Republicans saw it as factual”); Elizabeth Kolbert, *Why Facts Don’t Change Our Minds*, NEW YORKER (Feb. 27, 2017), <https://www.newyorker.com/magazine/2017/02/27/why-facts-dont-change-our-minds>.

22. Fazio et al., *supra* note 21, at 993.

23. See generally *id.* at 993–1002.

24. Brian Resnick, *Alex Jones and the Illusory Truth Effect, Explained*, VOX (June 19, 2017), <https://www.vox.com/science-and-health/2017/6/17/15817056/alex-jones-megyn-kelly-lies-nbc-psychology-illusory-truth>.

25. See Harwell, *supra* note 10 (“All that’s needed to make a persuasive mimicry within a matter of hours is a computer and a robust collection of photos, such as those posted by the millions onto social media every day.”).

26. See Ari Ezra Waldman, *A Breach of Trust: Fighting Nonconsensual Pornography*, 102 IOWA L. REV. 709, 716–19 (2017).

discourse.²⁷ Trust gives us the confidence to share personal information with others, from lawyers to loved ones.²⁸ When trust is strong and buttressed by law, social relationships are strong, which makes society strong, as well. But when any extant image can be weaponized to harm, whether as fodder for blackmail, revenge, or ongoing harassment, the safety and security society requires disappears. The threat of exploitation leads to silencing and self-censorship.²⁹

All of these concerns are exacerbated when marginalized groups are disproportionately targeted for abuse. As with other forms of nonconsensual pornography, as well as most other forms of technology-facilitated harassment, the costs of deep fakes are not equally borne across society. Research demonstrates that women and racial and sexual minorities are more likely to be victimized by these abuses and to suffer more severe consequences because of them.³⁰ High-profile cases of deep-fake pornography so far have almost exclusively involved women, particularly celebrity women.³¹ And the technology is not limited to sexual representations. The technology can just as easily be harnessed to sow political and social discord.³² It is not hard to imagine deep-fake videos of racial minorities committing crimes or LGBT individuals abusing children, to say nothing of audiovisual material that could jeopardize national security or stoke global conflict. The danger is particularly acute where distrust of certain individuals or communities already exists: A deep-fake video can leverage our confirmation biases to edge its way into the national discourse.³³

And yet, when confronted with legislation to criminalize revenge porn or proposals to constrain deep fakes, the civil libertarian position is often that it is better to leave these forms of speech unregulated, because even well-intentioned limitations on speech, especially on the internet, will ultimately harm the most vulnerable in the end.³⁴

27. See Chesney & Citron, *supra* note 7 (manuscript at 23) (discussing how deep-fake technology has the potential to erode trust in institutions).

28. See generally Ari Ezra Waldman, *Safe Social Spaces*, 96 WASH. U. L. REV. (forthcoming 2019).

29. See Citron & Franks, *supra* note 12, at 352–53.

30. See *supra* notes 14–15.

31. See Harwell, *supra* note 10.

32. See Chesney & Citron, *supra* note 7 (manuscript at 23–24).

33. See, e.g., Alessandro Bessi et al., *Science vs Conspiracy: Collective Narratives in the Age of Misinformation*, PLOS ONE, Feb. 2015, at 9, <http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0118093&type=printable> (concluding that polarized communities rise up around conspiracy theory posts on Facebook); Delia Mocanu et al., *Collective Attention in the Age of (Mis)Information*, 51 COMPUTERS HUM. BEHAV. 1198, 1199 (2015) (finding that while “normative social influence” and “the coherence with the [consumer’s] individual system of beliefs” affected “[t]he process of acceptance of a claim,” a claim’s origin or whether it was substantiated did not).

34. See STROSSEN, *supra* note 2, at 81; Gale & Strossen, *supra* note 5, at 180.

There are two premises, neither of which withstands scrutiny, often at work in this position. The first foundation is history. Since the Alien and Sedition Acts,³⁵ legislatures have indeed enacted laws that disproportionately targeted minority and dissident voices. Public schools have used their power to silence queer students³⁶ and anti-war protestors,³⁷ among many other minority voices.³⁸ And appeals to community morality allowed state actors to silence pro-equality discourse and information necessary for the safety and solidarity of the queer community.³⁹ It does not follow from these uncomfortable realities, however, that it is impossible to create well-crafted policies that criminalize revenge porn or harmful deep fakes. Arguing that we should not enact reasonable limitations on harmful speech because historical speech restrictions often targeted minority voices is like saying we should not criminalize rape because the criminal law has long been used to subjugate women.⁴⁰

The second premise upon which the civil libertarian position relies is power asymmetry. Restrictions on speech, the argument goes, will always end up favoring the powerful because the powerful can more easily avoid regulations or change them in their interests in the political sphere. But this asymmetry is at work as much in the absence of regulation as in the presence of it. Majorities, particularly online, can leverage their power to harass, target, and extort minority voices into silence. Without legal support, women and other minorities will find themselves hounded out of online social

35. Sedition Acts, ch. 74, 1 Stat. 596 (1798); ch. 75, 40 Stat. 553 (1918).

36. See, e.g., HUMAN RIGHTS WATCH, "LIKE WALKING THROUGH A HAILSTORM": DISCRIMINATION AGAINST LGBT YOUTH IN US SCHOOLS (2016), <https://www.hrw.org/report/2016/12/07/walking-through-hailstorm/discrimination-against-lgbt-youth-us-schools>.

37. *Tinker v. Des Moines Indep. Cmty. Sch. Dist.*, 393 U.S. 503 (1969) (finding unconstitutional a school district's policy that prohibited students from wearing black armbands to protest the Vietnam War).

38. See e.g., Salvador Hernandez & Brianna Sacks, *Here's How Some Schools Punished Students for Walking out to Protest Gun Violence*, BUZZFEED NEWS (Mar. 15, 2018), <https://www.buzzfeednews.com/article/salvadorhernandez/how-schools-punished-students-for-gun-walkouts> (punishing students for walkouts against gun violence); Kristine Phillips, *Florida Sixth-Grader Arrested after Dispute with Teacher over Pledge of Allegiance*, WASH. POST (Feb. 18, 2019), https://www.washingtonpost.com/education/2019/02/17/florida-sixth-grader-charged-with-misdemeanor-after-refusing-recite-pledge-allegiance/?utm_term=.358f2df37904 (arresting student after telling teacher the national anthem and flag were racist); Eric Levitz, *Red States Defend Free Speech—With Laws Punishing Campus Protest*, NEW YORK: INTELLIGENCER (June 14, 2018), <https://nymag.com/intelligencer/2018/06/red-states-defend-free-speech-with-laws-punishing-protesters.html> (passing "Goldwater" bills to impose punishments on students for engaging in protest activity).

39. *One, Inc. v. Olesen*, 355 U.S. 371 (1958) (invalidating the Comstock Laws that allowed the federal government to stop the transmission of queer-related literature deemed "immoral" through the mail).

40. See, e.g., GAYLE RUBIN, *DEVIATIONS: A GAYLE RUBIN READER* (2011) (discussing the ways the United States has criminalized sex and sexuality, discriminating against women and queer persons).

spaces.⁴¹ Majorities have become very effective at leveraging their unrestricted megaphones to their benefit.

Though civil libertarianism is often associated with liberal, progressive values, kneejerk resistance to regulation is inherently a conservative position that valorizes the status quo. The problem of deep fakes exposes the limitations of conventional free speech discourse and the civil libertarian position. When the free speech status quo is dominated by fraudulent, unanswerable, oppressive speech, failing to act means siding with the oppressor.

41. See Catherine Piner, *Feminist Writer Jessica Valenti Takes a Break from Social Media After Threat Against Her Daughter*, SLATE (July 28, 2016 5:01 PM), http://www.slate.com/blogs/xx_factor/2016/07/28/feminist_writer_jessica_valenti_takes_a_break_from_social_media_after_threat.html.