# FAKE

*Andrea M. Matwyshyn*[†] *& Miranda Mowbray*[††]

*The Internet today is full of fake people and fake information. Trust in both technology and institutions is in a downward spiral. This Article offers a novel comprehensive framework for calibrating a legal response to technology "fakery" through the lens of information security. Introducing the problems of Internet "MIST"—manipulation, impersonation, sequestering, and toxicity—it argues that these MIST challenges threaten the future viability of the Internet through two morphed dynamics destructive to trust. First, the arrival of the Internet-enabled "long con" has combined traditional con artistry with enhanced technological capability for data collection. Second, the risk of a new "PSYOP industrial complex" now looms. This chimera fuses techniques and employees from military psychological operations with marketing technologies of message hyper-personalization in order to target civilian audiences for behavioral modification. To address these two problematic dynamics through law, the Article constructs a broader theory of Internet untrustworthiness and fakery regulation.*

*Legal scholarship currently conflates two materially different forms of trust—trust in code and trust in people. As such, first, the Article imports the distinction from computer science theory between "trusted" and "trustworthy" systems. Next, engaging with the work of marketing theorist Edward Bernays, philosopher Jacques Ellul, and the theory of illusionism, this Article explains that determinations of intent/knowledge and context can serve as guideposts for legal paradigms of "untrustworthiness." This recognition offers a path forward for legal analysis of technology fakery and Internet harms. By engaging with multiple threads of First Amendment jurisprudence and scholarship, the Article next sketches the First Amendment bounds for regulation of untrustworthy technology content and conduct. Finally, this Article presents a novel framework inspired by the philosophy*

† Associate Dean of Innovation and Professor of Law and Engineering Policy, Penn State Law (University Park); Professor of Engineering Design, Penn State Engineering; Founding Director Penn State Policy Innovation Lab of Tomorrow (PILOT); Founding Director, Penn State Law Manglona Lab for Gender and Economic Equity. The authors wish to thank Michael Antonino, Gaia Bernstein, Matt Blaze, Leisel Bogan, Julie Cohen, Lilian Edwards, Mark Geistfeld, Sue Glueck, Tatters Glueck, Gray C. Hopper, F. Scott Kieff, Christopher Marsden, Gaela Normile, Martin Redish, Daniel Susser, Marcia Tiersky, and Jessica Wilkerson.

†† Honorary Lecturer, Centre for Doctoral Training in Interactive AI, University of Bristol; Academic Affiliate, Penn State Policy Innovation Lab of Tomorrow (PILOT).

*of deception to frame future legal discussions of untrustworthy technology fakery—the NICE framework. NICE involves an evaluation of three variables—the legal nature of the technology content or conduct, the intent and knowledge of the faker, and the sensitivity of the context. Thus, NICE leverages traditional legal models of regulation to the greatest extent possible in addressing technology fakery. This Article concludes with examples of regulatory approaches to technology fakery informed by the NICE framework.*

TABLE OF CONTENTS

## INTRODUCTION

> *With traps and obstacles and hazards confronting us on every hand, only blindness or indifference will fail to turn in all humility, for guidance or for warning, to the study of examples.*
>
> *—Justice Benjamin N. Cardozo*[1]

Ours is the age of the Internet clapback and the competing dank meme. But, despite our intuitions of novelty, our current struggles with information integrity and technology are merely the latest round of a recurring historical contest.[2] We have always been at war with technology fakery.

---

1 BENJAMIN N. CARDOZO, *Law and Literature*, *in* LAW AND LITERATURE AND OTHER ESSAYS AND ADDRESSES 3, 9 (1931).

2 New technologies and information fakery have regularly upset our national discourse. *See, e.g.*, JOANNE B. FREEMAN, THE FIELD OF BLOOD: VIOLENCE IN CONGRESS AND THE ROAD TO CIVIL WAR 169–71 (2018).

In the 1830s, the arrival and mass adoption[3] of the telegraph brought with it the rise of conspiracy theories,[4] even disrupting the dynamics of congressional constituent communications.[5] As argued by Professor Joanne Freeman, "[t]he telegraph was the social media of its day," and it "spread journalistic hot-takes throughout the nation with greater reach and speed than ever before."[6] For example, partially because of the telegraph,[7] members of Congress no longer controlled the reach, speed, or integrity of the information around their affairs of honor,[8] and political upheaval further fomented as a result.[9]

Parallel dynamics repeated with the arrival of radio and television in the United States. In 1907, key breakthroughs in amplification tube technology paved the way for the first image to be instantaneously transmitted through telegraph wires.[10] This shift also enabled the arrival of radio and the first moving images transmitted by television in the 1920s.[11] By the 1930s, radio had reached broad public adoption, and fake information, conspiracy theories, and the "snake oil" salespeople[12] of earlier eras took to the airwaves to perpetrate fraud.[13] Innocent

---

3 By 1845, the telegraph was already in use to assist in criminal apprehensions. *See* Thomas McMullan, *The World's First Hack: The Telegraph and the Invention of Privacy*, GUARDIAN (July 15, 2015, 3:00 AM), https://www.theguardian.com/technology/2015/jul/15/first-hack-telegraph-invention-privacy-gchq-nsa [https://perma.cc/653S-E588].

4 As explained by Professor Joanne Freeman, "The public [was] learning all kinds of things from all kinds of people and they [could not] tell what's true and false. Conspiracy theories start[ed] to spread because of the confusion. . . . [T]he telegraph did what social media does today." *Politics of American Dueling*, at 34:55 (C-SPAN television broadcast Feb. 20, 2020), https://www.c-span.org/video?469539-1/politics-american-dueling [https://perma.cc/RB28-YC2B]; *see also* FREEMAN, *supra* note 2.

5 FREEMAN, *supra* note 2.

6 Joanne B. Freeman, *Why 1850 Doesn't Feel So Far Away*, N.Y. TIMES (Jan. 29, 2021), https://www.nytimes.com/2021/01/29/opinion/political-violence-congress.html [https://perma.cc/88CV-4ZQD].

7 *Id.*

8 *See Politics of American Dueling*, *supra* note 4, at 34:06.

9 *Id.* at 34:55.

10 *Sending Photographs by Telegraph: Professor Korn Has Triumphantly Succeeded in Transmitting Portraits over Long Distances by Wire—Experiments in France and Germany Conclusive—Description of the Marvelous Instrument*, N.Y. TIMES, Feb. 24, 1907 (pt. 3), at 7, https://www.nytimes.com/1907/02/24/archives/sending-photographs-by-telegraph-professor-korn-has-triumphantly.html [https://perma.cc/J75U-9T2W].

11 *Id.*; *see also Audion*, BRITANNICA, https://www.britannica.com/technology/Audion [https://perma.cc/99B3-CKK8]; *1920s–1960s: Television*, ELON UNIV., https://www.elon.edu/u/imagining/time-capsule/150-years/back-1920-1960 [https://perma.cc/2BS8-752C].

12 By the 1950s, public awareness of the risks of "snake oil salesmen" and their fakery in mass media was widespread. For example, the 1950s television series *Trackdown* included an episode about a snake oil salesman named Trump who claimed that only he could save the town by building a wall around it. Dan Evon, *Did a 1950s TV Episode Feature a Character Named Trump Who Offered to Build a Protective Wall?*, SNOPES (Jan. 9, 2017), https://www.snopes.com/fact-check/trackdown-trump-character-wall [https://perma.cc/PRA6-AMVT].

13 *See* discussion *infra* Part II.

confusion also existed: in 1938, a broadcast work of science fiction, *The War of the Worlds*, caused confusion when listeners tuned in to hear a fake journalist seemingly provide real-time updates about an alleged alien invasion of Earth.[14] Smallpox and polio vaccine disinformation and misinformation in print predated COVID-19 vaccine disinformation and misinformation online.[15]

We too live in an era where a new technology has amplified the reach of dueling words and images and where fakery happens in both real time and on a time-shifted basis. Speed and amplification of information—both fake and real—have increased as computing power has increased.[16] And just as in prior eras, the risks of corrupted information damaging public discourse loom large.[17] Internet fakery has already impacted the operation of our markets,[18] our republic's governance and national security,[19] and the public's sense of trust in the Internet and institutions more generally.[20]

---

14 *See* A. BRAD SCHWARTZ, BROADCAST HYSTERIA: ORSON WELLES'S *WAR OF THE WORLDS* AND THE ART OF FAKE NEWS 3–4 (2015); Carl Holm, *The Radio Drama that Shocked America 80 Years Ago and the Modern Birth of Fake News*, DW (Oct. 26, 2018), https://www.dw.com/en/the-radio-drama-that-shocked-america-80-years-ago-and-the-modern-birth-of-fake-news/a-46052965 [https://perma.cc/F5AB-5PD3].

15 *See, e.g.*, Lawrence O. Gostin, Op-Ed, *Global Polio Eradication: Espionage, Disinformation, and the Politics of Vaccination*, 92 MILBANK Q. 413, 414 (2014); Dan MacGuill, *Did a 1930s Cartoon Warn of Vaccine Misinformation?*, SNOPES (May 13, 2021), https://www.snopes.com/fact-check/1930s-cartoon-vaccine-warning [https://perma.cc/XL4X-4LXW]; Alexandra Lord, *Anti-Vaccination in America*, NAT'L MUSEM OF AM. HIST. (Aug. 31, 2015), https://americanhistory.si.edu/blog/anti-vaccination-america [https://perma.cc/4Y63-Q8YU].

16 The seeming permanence of Internet "memory" has galvanized "right to forget" legislation attempts in the EU and calls for modified versions of that right in the United States. *See, e.g.*, Andrea M. Matwyshyn, *Generation C: Childhood, Code, and Creativity*, 87 NOTRE DAME L. REV. 1979 (2012) (arguing in favor of applying minority doctrine to end user license agreements and privacy policies); *see also, e.g.*, Lilian Edwards & Edina Harbinja, *Protecting Post-Mortem Privacy: Reconsidering the Privacy Interests of the Deceased in a Digital World*, 32 CARDOZO ARTS & ENT. L.J. 83 (2013); Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149 (2018); VIKTOR MAYER-SCHÖNBERGER, DELETE: THE VIRTUE OF FORGETTING IN THE DIGITAL AGE (2009); Jeffrey Rosen, *The Right to Be Forgotten*, 64 STAN. L. REV. ONLINE 88 (2012).

17 *See infra* Section I.C.

18 *See* COMMODITY FUTURES TRADING COMM'N, CUSTOMER ADVISORY: BEWARE VIRTUAL CURRENCY PUMP-AND-DUMP SCHEMES (2018), https://cftc.gov/sites/default/files/2019-12/customeradvisory_pumpdump0218.pdf [https://perma.cc/DAY8-J884].

19 *See Russian Interference in 2016 U.S. Elections*, FBI, https://www.fbi.gov/wanted/cyber/russian-interference-in-2016-u-s-elections [https://perma.cc/HH45-WHD6].

20 *Survey Says People Don't Trust the Internet, What Needs to Change?*, UNITED NATIONS CONF. ON TRADE & DEV. (June 11, 2019), https://unctad.org/news/survey-says-people-dont-trust-Internet-what-needs-change [https://perma.cc/U7GX-JHCW]; Lee Rainie & Janna Anderson, *The Fate of Online Trust in the Next Decade*, PEW RSCH. CTR. (Aug. 10, 2017), https://www.pewresearch.org/internet/2017/08/10/the-fate-of-online-trust-in-the-next-decade [https://perma.cc/9L39-G85K].

No comprehensive theory of technology "fakery" currently exists in legal scholarship that effectively merges traditional and modern jurisprudence. As such, this Article offers a novel comprehensive legal framework to conceptualize fakery in technology contexts as an information security problem. Part I introduces the problem of Internet "MIST" or, more colloquially, the "Four Tarantulas of the Fakopalypse"—a modern variant of the 1990s meme of the Four Horsemen of the Infocalypse. It argues that four "tarantulas" of fakeness—manipulation, impersonation, sequestering, and toxicity—create a "MIST" of Internet fakery that harms trust. Part I then explains that two key dynamics complicate today's Internet fakery problems: the arrival of the Internet "long con" and the risk of creating a "PSYOP industrial complex." Using classical con-artistry theory, Part I first explains that the Internet "long con" merges timeless techniques of con artists with new information exploitation opportunities to more efficiently identify and exploit "suckers" or "marks." Second, using theory of propaganda and psychological operations (PSYOP), Part I explains that the techniques of Internet marketing have begun to blend with techniques and personnel from military psychological operations. This merger risks the rise of a "PSYOP industrial complex"—an enterprise that targets audiences for fakery and behavioral modification with progressively greater precision. In other words, psychological manipulation techniques learned from militarized operations in warfare appear to be transferring into use on civilian populations. These two dynamics have been instrumental in the progressive erosion of trust visible on the Internet today. Gleaning insights from marketing theory of Edward Bernays, the work of lawyer and philosopher of technology, Jacques Ellul, and the theory of illusionism, Part I concludes by reframing these dynamics, highlighting two elements that have historically underlain the assessment of untrustworthy content or conduct: the intent and knowledge of the faker and the context of the technology fakery.

Part II begins to map viable paths forward to halt Internet trust erosion due to fakery. It first introduces a core definitional distinction from computer science that is currently absent from the legal literature's discussion of technology fakery—the critical distinction between being trusted and being trustworthy. It argues that legal approaches to fakery should focus on the regulation of untrustworthiness. Turning next to the philosophy of trust, Part II differentiates assessment criteria for trustworthy versus untrustworthy code, objects, and people. Finally, Part II concludes by engaging with multiple threads of First Amendment scholarship and case law to identify the First Amendment limits for regulation of untrustworthy Internet content and conduct.

Part III then merges the insights from prior Parts to offer a reframe for legal discussions of untrustworthy Internet content. This new framework, NICE, involves an examination of three variables—the legal

nature of the fakery, the intent and knowledge of the faker, and the sensitivity of the context into which the fakery was injected. Part III then offers examples of regulatory approaches informed by the NICE framework for each category of Internet fakery/MIST.

## I. THE INTERNET WINTER OF OUR (DIS)CONTENT: THE TANGLED WEB OF TRUST

> *And thus we arrive at Lucian's weakness[,] . . . . a misguided admiration of the truth . . . . [and] those spiders, of mighty bigness, every one of which exceeded in size an isle of the Cyclades. "These were appointed to spin a web in the air between the Moon and the Morning Star, which was done in an instant, and made a plain champaign, upon which the foot forces were planted." Truly a very Colossus of falsehood . . . .*[21]

The earliest known work of science fiction in Western literature was a satirical novel.[22] Written in the second century, *Lucian's True History* by Lucian of Samosata chronicles the adventures of Lucian and the people of Earth as they battle invading attackers[23] and giant spiders with a country-sized[24] web.[25] Now, we, the people of Earth of the twenty-first century, face our own epic battle against our own Colossus of falsehood with a web in the air[26]—technology fakery.

Much has been written about the various legal challenges presented by fake Internet content and conduct. This ample body of existing scholarship has generally approached the dynamics of various forms of technology "fakery" in a segmented manner. The most ambitious scholarship has primarily focused on either the private sector dynamics of the Internet economy and its relationship to data[27] on the one hand and the political impact of propaganda and attempted election manipulation on the other.[28] Professor Julie Cohen presents an insightful discussion of political economy, applying perspectives of Karl Polanyi on the overall dynamics of the mutual constitution of the

---

21 Charles Whibley, *Introduction* to LUCIAN OF SAMOSATA, LUCIAN'S TRUE HISTORY (Francis Hickes trans., Project Gutenberg 2014) (1894).

22 *See* S.C. Fredericks, *Lucian's* True History *as SF*, SCI. FICTION STUD. (Mar. 1976), https://www.depauw.edu/sfs/backissues/8/fredericks8art.htm [https://perma.cc/XY5H-UKQV].

23 *See* LUCIAN OF SAMOSATA, *supra* note 21.

24 *Id.*

25 *Id.*

26 *Cf. id.*

27 SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

28 YOCHAI BENKLER, ROBERT FARIS & HAL ROBERTS, NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS (2018).

technology economy and its relationship to law.[29] Professor Cohen highlights the appropriation of intangible resources and "performative enclosure" driven by repetition,[30] as it connects with commodification/datafication; thus, her work reframes the response in law, particularly as technology platforms become increasingly prevalent.[31] By comparison, Professor Shoshana Zuboff offers a critique of the business dynamics of the data economy through the lens of what she terms Internet "surveillance capitalism,"[32] arguing that technology-mediated surveillance is an "instrumentarian power that fills the [trust] void, substituting machines for social relations, which amounts to *the substitution of certainty for society*."[33] However, both authors stop short of connecting their analyses with broader information security or national security concerns.[34] Meanwhile, Professors Yochai Benkler, Robert Faris, and Hal Roberts present a thoughtful analysis of the role of fake information in the 2016 U.S. presidential election.[35] In particular, the authors assert that they "believe there is an advantage to keeping separate the domain of politics, with its normative commitment to democracy, from the domain of commerce, and its normative commitment to welfare, consumer sovereignty, and consumer protection."[36] This Article builds on these scholars' noteworthy prior work. However, it adopts an intentionally contrary approach.

This Article instead expressly merges the dynamics that have been previously differentiated by other scholars: it presents an analysis that assumes the interweaving of the political/national security domain and the commercial one in its approach to technology fakery. Specifically, this Article explains that these two domains—commercial and political fakery—have always been functionally interwoven, as a matter of both technological and historical practice. Further, as Part II explains, they are converging as a matter of First Amendment jurisprudence. Thus, we argue, a single broader framework of technology fakery is needed for

---

29 Julie Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism*, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y (Nov. 21, 2019), https://cyber.harvard.edu/events/between-truth-and-power-legal-constructions-informational-capitalism [https://perma.cc/2MBW-KZR4].

30 *Id.*; The Berkman Klein Ctr. for Internet & Soc'y, *Between Truth and Power: Featuring Julie Cohen*, YOUTUBE, at 13:19 (Dec. 13, 2019), https://youtu.be/reH9g9PnA_4 [https://perma.cc/ML8K-N72S].

31 The Berkman Klein Ctr. for Internet & Soc'y, *supra* note 30, at 13:57.

32 The analysis presents an Internet-exceptionalist frame, and Zuboff stops short of connecting these dynamics with their legal, information security, and national security implications—another difference with this Article.

33 *See* ZUBOFF, *supra* note 27, at 384.

34 The authors stop short of fully exploring the history of exploitation of information to psychologically target populations and the role of experienced personnel in this enterprise.

35 *See* BENKLER, FARIS & ROBERTS, *supra* note 28.

36 *Id*. at 30.

combating Internet fakery. We offer one such approach, the NICE framework, in Part III.

But before we arrive at concrete proposals in Part III, we must start our discussion with an articulation of the meaning of the term "Internet fakery." The Section that follows defines "Internet fakery" as the problems of "MIST"—manipulation, impersonation, sequestration, and toxicity—and explores them through the metaphor of tarantulas—a colloquial homage not only to Lucian and his epic second-century–science-fiction spiders but also to early Internet history.

In one of the earliest articulations of Internet harms, in 1998, engineer and author Tim May argued that regulation of the Internet would be driven by a fear of the "specter of crypto anarchy" and the availability of technology tools that facilitate anonymous communication.[37] In particular, these new tools, argued May, would likely raise concerns from law enforcement over stopping the "Four Horsemen of the Infocalypse"—terrorism, child exploitation, drug dealing, and money laundering/organized crime.[38] While intended by May as tongue-in-cheek-hyperbolic commentary, these observations were largely accurate in predicting the first two decades of Internet law.[39] Combining Lucian of Samosata's giant spiders with May's Four Horsemen of the Infocalypse, we might colloquially recast our problem of Internet fakery and MIST—manipulation, impersonation, sequestering, and toxicity—as the Four Tarantulas of the Fakopalypse.

---

37 Timothy C. May, *The Crypto Anarchist Manifesto*, SPUNK LIBR., http://www.spunk.org/library/comms/sp000151.html [https://perma.cc/SR2M-8MTR] ("A specter is haunting the modern world, the specter of crypto anarchy. Computer technology is on the verge of providing the ability for individuals and groups to communicate and interact with each other in a totally anonymous manner.").

38 "How will privacy and anonymity be attacked?"

> —the downsides just listed are often cited as a reason we can't have "anonymity"
>
> —like so many other "computer hacker" items, as a tool for the "Four Horsemen": drug-dealers, money-launderers, terrorists, and pedophiles.

TIMOTHY C. MAY, *Anonymity, Digital Mixes, and Remailers*, *in* THE CYPHERNOMICON: CYPHERPUNKS FAQ AND MORE 8.3.4 (1994), https://groups.csail.mit.edu/mac/classes/6.805/articles/crypto/cypherpunks/cyphernomicon/CP-FAQ [https://perma.cc/HJD2-B29J].

39 May's original observations were made in the context of anonymization tools; however he also used the phrase "Four Horsemen of the Infocalypse" later on in the document in the more general context of the "Net of the Future." *Id.* at 17.5.7. See also the biblical concept of the four horsemen of the apocalypse. *Four Horsemen of the Apocalypse*, BRITANNICA, https://www.britannica.com/topic/four-horsemen-of-the-Apocalypse [https://perma.cc/3U2N-6KP4].

### A. *The Problem of Internet MIST (or The Four "Tarantulas" of the Fakopalypse)*

In the mid-1990s, the Silicon Valley offices of AT&T suffered from a physical security (literal) "bug" race condition[40]: the company's basement was overrun by a colony of tarantula spiders.[41] Mostly unbothered in their daily routines, AT&T employees (and the spiders) continued their work unflummoxed; the employees simply chose to cede operational control of the basement. As recounted by former employees, they avoided entering the new "tarantula cave" and just hoped for the best.[42] But, unfortunately for the AT&T employees, their tarantula detente was destined for failure: tarantulas tend to mate rather efficiently,[43] and they demonstrate comparative longevity, with females living up to thirty years.[44] It was likely only a matter of time before the tarantulas expanded their territory into the workspace inhabited by humans.[45] Yet, despite this looming spider invasion afoot (and underfoot), superficially the office appeared to be working as usual.[46]

While the last two decades have been characterized by restrained Internet regulatory action, staying this course places us and the future of the Internet, at best, in the position of the AT&T employees—in an unsustainable detente.[47] We have reached the equivalent of the "tarantula cave" era of technology fakery. Much like the specter of an impending arachnid incursion,[48] deficits of user trust loom large over

---

40 A race condition exists when two separate processes "that would access a shared resource could do so in such a way as to cause unexpected results." *See, e.g.*, privatehuff, Comment to *What Is a Race Condition?*, STACK OVERFLOW (Aug. 29, 2008, 5:01 PM), https://stackoverflow.com/questions/34510/what-is-a-race-condition [https://perma.cc/5D5W-95Q8]; *see also CWE-362: Concurrent Execution Using Shared Resource with Improper Synchronization ("Race Condition")*, COMMON WEAKNESS ENUMERATION, http://cwe.mitre.org/data/definitions/362.html [https://perma.cc/4HPL-S9DX].

41 E-mail from Matt Blaze, Professor, Georgetown L., to Andrea M. Matwyshyn (Mar. 1, 2020) (on file with author).

42 *Id.*

43 *Tarantulas*, NAT'L GEOGRAPHIC, https://www.nationalgeographic.com/animals/invertebrates/group/tarantulas [https://perma.cc/RMX2-JVME].

44 *Id.*

45 *See id.*

46 E-mail from Matt Blaze to Andrea M. Matwyshyn, *supra* note 41.

47 And, unlike the basement of an office building, the Internet does not offer physical barriers to impede the "tarantulas" of Internet fakery from spreading to new spaces.

48 *See supra* notes 21–26 and accompanying text.

the future of the Internet. Fake news,[49] fake videos,[50] fake users,[51] fake reviews,[52] fake websites,[53] and fake services[54] with fake offers,[55] fake algorithms,[56] and fake results[57] seem omnipresent on the Internet. As explained by a recent study from the Pew Internet and American Life Project, "Trust has not been having a good run in recent years, and there is considerable concern that people's uses of the Internet are a major contributor to the problem. For starters, the Internet was not designed with security protections or trust problems in mind."[58] Seventy-nine percent of Americans say they are not confident that companies will admit mistakes and take responsibility for misuse or compromise of information.[59] Similarly, "[eighty-one percent] of the public say that the potential risks they face because of data collection by companies outweigh the benefits."[60] These drops in Internet trust also coincide with eroding trust in institutions and government generally,[61] as fake

---

49 Rory Cellan-Jones, *Coronavirus: Fake News Is Spreading Fast*, BBC (Feb. 26, 2020), https://www.bbc.com/news/technology-51646309 [https://perma.cc/JM37-Z7FW].

50 Donie O'Sullivan, *When Seeing Is No Longer Believing*, CNN, https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes [https://perma.cc/M9XV-BEQQ].

51 Alan Wolk, *Facebook's Got Billions of Fake Users, Though that May Be the Least of Their Problems*, FORBES (June 29, 2018, 9:24 PM), https://www.forbes.com/sites/alanwolk/2018/06/29/facebooks-got-billions-of-fake-users-though-that-may-be-the-least-of-their-problems [https://perma.cc/CPS4-5ZYQ].

52 Rick Broida, *How to Spot Fake Reviews on Amazon, Best Buy, Walmart, and Other Sites*, CNET (Mar. 4, 2019, 8:41 AM), https://www.cnet.com/how-to/spot-fake-reviews-amazon-best-buy-walmart [https://perma.cc/BL7E-SAU5].

53 *How to Spot a Fake Web Site and Not Get Phished*, AT&T, https://www.att.com/support/article/smb-Internet/KM1188156 [https://perma.cc/BH8E-4RE9].

54 *Internet Auction Fraud*, FBI, https://www.fbi.gov/scams-and-safety/common-scams-and-crimes/internet-auction-fraud [https://perma.cc/Z2JR-ZUQ3].

55 *Investment Fraud*, FBI, https://www.fbi.gov/scams-and-safety/common-fraud-schemes/investment-fraud [https://perma.cc/3A93-NSFH].

56 Tarry Singh, *Struggling with Fake AI? Here's How to Become a Real AI Company*, FORBES (Feb. 28, 2020, 2:36 AM), https://www.forbes.com/sites/cognitiveworld/2020/02/28/struggling-with-fake-ai-heres-how-to-become-a-real-ai-company [https://perma.cc/AN42-KMX4].

57 Erin Griffith, *Theranos and Silicon Valley's "Fake It till You Make It" Culture*, WIRED (Mar. 14, 2018, 3:12 PM), https://www.wired.com/story/theranos-and-silicon-valleys-fake-it-till-you-make-it-culture [https://perma.cc/M85T-HCVX].

58 Rainie & Anderson, *supra* note 20.

59 BROOKE AUXIER ET AL., PEW RSCH. CTR., AMERICANS AND PRIVACY: CONCERNED, CONFUSED AND FEELING LACK OF CONTROL OVER THEIR PERSONAL INFORMATION (2019), https://www.pewresearch.org/Internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information [https://perma.cc/BYD4-UDMN].

60 *Id.*

61 According to the Edelman Trust Barometer, 2017 was the first time the study found a decline in trust across business, media, government, and NGOs. Matthew Harrington, *Survey: People's Trust Has Declined in Business, Media, Government, and NGOs*, HARV. BUS. REV. (Jan.

Internet content helps exacerbate a self-reinforcing trust spiral downward.

At least four kinds of problematic technology "fakery" are visible today—the problems of manipulation, impersonation, sequestering, and toxicity (MIST).

### 1. The Problem of Manipulation

Nearly three decades ago, a famous *New Yorker* cartoon once announced, "On the Internet, nobody knows you're a dog."[62] Today, the ability to track individual users online has increased, as has the technological ability for companies and national security experts to attribute speech and conduct.[63] However, the challenges faced by average users in parsing whether to trust content and persons they encounter on the Internet have arguably become even more formidable. In particular, users regularly encounter content that has been manipulated, either in substance or in its superficial attribution.[64]

Borrowing the lens of computer security, the challenges of fake and manipulated technology content might be reframed more elegantly and simply as (the latest iteration of) problems of information integrity in the technical sense.[65] These problems of manipulation might be divided into two categories—manipulation of content and manipulation of authenticity.[66]

---

16, 2017), https://hbr.org/2017/01/survey-peoples-trust-has-declined-in-business-media-government-and-ngos [https://perma.cc/4NMR-UEWR]; PEW RSCH. CTR., PUBLIC TRUST IN GOVERNMENT REMAINS NEAR HISTORIC LOWS AS PARTISAN ATTITUDES SHIFT (2017), https://www.people-press.org/2017/05/03/public-trust-in-government-remains-near-historic-lows-as-partisan-attitudes-shift [https://perma.cc/489Y-KY2W].

62 Michael Cavna, *"NOBODY KNOWS YOU'RE A DOG": As Iconic Internet Cartoon Turns 20, Creator Peter Steiner Knows the Joke Rings As Relevant As Ever*, WASH. POST (July 31, 2013), https://www.washingtonpost.com/blogs/comic-riffs/post/nobody-knows-youre-a-dog-as-iconic-internet-cartoon-turns-20-creator-peter-steiner-knows-the-joke-rings-as-relevant-as-ever/2013/07/31/73372600-f98d-11e2-8e84-c56731a202fb_blog.html [https://perma.cc/3L6U-AS2S].

63 *See, e.g.*, Herbert Lin, *Attribution of Malicious Cyber Incidents* (Hoover Inst. Aegis Paper Series, No. 1607, 2016), https://www.hoover.org/sites/default/files/research/docs/lin_webready.pdf [https://perma.cc/M424-YQBC].

64 Discussions around manipulative content and conduct may engage with terms such as disinformation, misinformation, propaganda, and other related, though arguably distinct, terminology. *See* Andrei Richter, *Fake News and Freedom of the Media*, 8 J. INT'L MEDIA & ENT. L. 1, 7–9 (2018).

65 NIST defines integrity as "[g]uarding against improper information modification or destruction, and includes ensuring information non-repudiation and authenticity." *Integrity*, NIST, https://csrc.nist.gov/glossary/term/integrity [https://perma.cc/RBL7-LNQ4].

66 Manipulation generally describes a form of influence that is neither coercion nor rational persuasion. In general, philosophers have distinguished between ordinary and global

### a. Manipulation of Content

The film *Catch Me if You Can*[67] describes the adventures of con artist Frank Abagnale Jr., a faker so skilled that allegedly a police chief once quipped: "Frank Abagnale could write a check on toilet paper, drawn on the Confederate States Treasury, sign it 'U.R. Hooked' and cash it at any bank in town, using a Hong Kong driver's license for identification."[68] According to the book of the same title by Stan Redding with Abagnale, Abagnale allegedly landed a job at the Louisiana State Attorney General's office after passing the bar exam but providing a forged law transcript from Harvard University.[69] As this situation of a potentially forged law transcript illustrates, the problem of fake information is not a new problem arising from the Internet.[70] While the law has perhaps most often analyzed the idea of manipulation in the context of document forgery[71] and market manipulation in securities regulation,[72] legal concern and scholarship over manipulation of information (under various definitions) has also arisen in the context

manipulation. While ordinary manipulation refers to acts that limit exercise of free will in a particular context, global manipulation is typically considered to deprive targets entirely of free will or autonomy. For example, in the context of medical ethics, discussion often centers on questions of whether consent is not manipulated. *See* RUTH R. FADEN, TOM L. BEAUCHAMP & NANCY M. P. KING, A HISTORY AND THEORY OF INFORMED CONSENT 54 (1986). For a different approach to manipulation in marketing, see Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1034 (2014) (arguing that manipulation by firms includes "purposefully render[ing] consumers vulnerable" through use of Big Data).

67 CATCH ME IF YOU CAN (Dreamworks Pictures 2002).

68 FRANK W. ABAGNALE & STAN REDDING, CATCH ME IF YOU CAN: THE TRUE STORY OF A REAL FAKE 116 (paperback ed. 2002).

69 *Id.* at 101–03.

70 Forgery of various types has long been a source of concern for law enforcement, economic institutions, and other high-credibility contexts. *See, e.g.*, Katelyn Polantz & Evan Perez, *Exclusive: Former FBI Lawyer Under Investigation After Allegedly Altering Document in 2016 Russia Probe*, CNN (Nov. 22, 2019, 12:36 PM), https://www.cnn.com/2019/11/21/politics/fbi-fisa-russia-investigation/index.html [https://perma.cc/4GJK-W82H].

71 As described by the *Harvard Law Review*, "[T]he examination of suspected and disputed documents has become a recognized scientific specialty." Albert S. Osborn, Book Review, 46 HARV. L. REV. 739, 739 (1933) (reviewing FRANK BREWESTER, CONTESTED DOCUMENTS AND FORGERIES (1932)).

72 *See, e.g.*, Lewis D. Lowenfels, *Sections 9(a)(1) and 9(a)(2) of the Securities Exchange Act of 1934: An Analysis of Two Important Anti-Manipulative Provisions Under the Federal Securities Laws*, 85 NW. U. L. REV. 698, 698 (1991).

of currency,[73] art forgery,[74] identity documentation,[75] and other legal situations.[76]

At least three types of manipulated Internet content concerns have received meaningful discussion in the legal scholarship—fake news,[77] fake videos, and fake reviews. Legal scholarship has addressed the intent of fake news creators[78] and its distortion of the electoral process,[79] and it has described fake news as leading to a "cascade of cynicism,"[80] a broader "distrust for all media production,"[81] and a distrust of expertise in general.[82] For example, Professor Alice Marwick has highlighted that the function of identity signaling often overrides concerns over accuracy for some participants in fake news purveyance,[83] complicating the viability of the two primary solutions proposed for combating fake news[84]—fact-checking and media literacy.[85] Additional scholars have highlighted the role of private versus public sector interventions in addressing manipulated content[86] and argued that current First

---

73 *See, e.g.*, Daniel C.K. Chow, *Can the United States Impose Trade Sanctions on China for Currency Manipulation?*, 16 WASH. U. GLOB. STUD. L. REV. 295, 295 (2017).

74 *See, e.g.*, Gregory Day, *Explaining the Art Market's Thefts, Frauds, and Forgeries (and Why the Art Market Does Not Seem to Care)*, 16 VAND. J. ENT. & TECH. L. 457, 457 (2014); Leila A. Amineddoleh, *Are You Faux Real? An Examination of Art Forgery and the Legal Tools Protecting Art Collectors*, 34 CARDOZO ARTS & ENT. L.J. 59, 61 (2016); Joseph C. Gioconda, *Can Intellectual Property Laws Stem the Rising Tide of Art Forgeries*, 31 HASTINGS COMMC'NS & ENT. L.J. 47, 49–50 (2008); Derek Fincham, *Authenticating Art by Valuing Art Experts*, 86 MISS. L.J. 567, 575 (2017); Michael J. Clark, *The Perfect Fake: Creativity, Forgery, Art and the Law*, 15 DEPAUL-LCA J. ART & ENT. L. 1, 5 (2004).

75 37 AM. JUR. 2D *Fraud and Deceit* § 50, Westlaw (database updated Aug. 2021).

76 *See, e.g.*, Case Note, *Wills—Olographic Will—Use of Typewriter*, 27 YALE L.J. 142 (1917).

77 "Nearly seven-in-ten (68%) say made-up news and information greatly affects Americans' confidence in government institutions, and roughly half (54%) say it is having a major impact on Americans' confidence in each other." Michael Dimock, *An Update on Our Research into Trust, Facts and Democracy*, PEW RSCH. CTR. (June 5, 2019), https://www.pewresearch.org/2019/06/05/an-update-on-our-research-into-trust-facts-and-democracy [https://perma.cc/CHM4-AXSV].

78 Jessica Pepp, Eliot Michaelson & Rachel Katharine Sterken, *What's New About Fake News?*, 16 J. ETHICS & SOC. PHIL. 67, 67–68 (2019).

79 Joel Timmer, *Fighting Falsity: Fake News, Facebook, and the First Amendment*, 35 CARDOZO ARTS & ENT. L.J. 669, 671 (2017).

80 Mark Verstraete & Derek E. Bambauer, *Ecosystem of Distrust*, 16 FIRST AMEND. L. REV. 129, 130 (2018).

81 *Id.*

82 *Id.* at 143.

83 Alice E. Marwick, *Why Do People Share Fake News? A Sociotechnical Model of Media Effects*, 2 GEO. L. TECH. REV. 474, 505 (2018).

84 *Id.* at 507.

85 Lili Levi, *Real "Fake News" and Fake "Fake News,"* 16 FIRST AMEND. L. REV. 232, 237–38 (2018) (arguing in favor of a focus on "platform self-regulation, audience information literacy, and empowerment of the press itself" to combat fake news).

86 *See, e.g.*, Richard L. Hasen, *Cheap Speech and What It Has Done (to American Democracy)*, 16 FIRST AMEND. L. REV. 200, 202 (2018).

Amendment doctrine appears to protect fake news.[87] As explained by two scholars, "The abundance of fake news is accompanied by claims that unfavorable but factual news is itself fake. By sowing seeds of distrust, false claims of fake news are designed to erode trust in the press . . . ."[88] Still, other scholars have raised concerns that the Supreme Court's commitment to the free flow of information under current First Amendment doctrine may limit the efficacy of policy interventions to limit fake news, even under the auspices of fraud prevention.[89] Further, the copyright status of fake news has also been the subject of some legal analysis,[90] raising the specter that the fictionalized nature of manipulated content might perversely grant it a form of potentially superior intellectual property status over truthful information.

Legal scholars have also devoted substantial thought to the context of fake and manipulated videos, such as "deep fakes." As recent press coverage explains, with the help of readily available video editing tools, video content may be faked or manipulated to create a false impression of the drunkenness of a public figure[91] or to time-shift content into a more desirable version of events.[92] As explained by Professors Robert Chesney and Danielle Citron, one of the risks represented by deep fakes arises from a form of "Truth Decay" and a "Liar's Dividend"—that "wrongdoers may find it easier to cast doubt on real recordings of their mischief."[93] Professors Mary Anne Franks and Ari Waldman argue that "deep fakes undermine free speech itself, at least of its targets. . . . [and] weaponize targets' speech against themselves, harvesting their photos, videos, and audio recordings to create increasingly realistic, fraudulent representations," and that "deep fakes erode the trust that is necessary for social relationships and political discourse."[94] Professor Nina Brown

---

87 *See* Clay Calvert, Stephanie McNeff, Austin Vining & Sebastian Zarate, *Fake News and the First Amendment: Reconciling a Disconnect Between Theory and Doctrine*, 86 U. CIN. L. REV. 99, 106 (2018).

88 Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 YALE J.L. & TECH. 106, 150 (2019).

89 For example, Professor Ari Waldman has argued that the marketplace of ideas analogy is a powerful barrier to implementation of most current fake news policy proposals. Ari Ezra Waldman, *The Marketplace of Fake News*, 20 U. PA. J. CONST. L. 845, 864–66 (2018).

90 Cathay Y. N. Smith, *Truth, Lies, and Copyright*, 20 NEV. L.J. 201, 220–21 (2019).

91 Maheen Sadiq, *Real v Fake: Debunking the "Drunk" Nancy Pelosi Footage—Video*, GUARDIAN (May 24, 2019, 12:38 PM), https://www.theguardian.com/us-news/video/2019/may/24/real-v-fake-debunking-the-drunk-nancy-pelosi-footage-video [https://perma.cc/GFP7-SYUY].

92 Alex Ward, *Mike Bloomberg Tweeted a Doctored Debate Video. Is It Political Spin or Disinformation?*, VOX (Feb. 20, 2020, 6:35 PM), https://www.vox.com/2020/2/20/21145926/mike-bloomberg-debate-video-twitter-fake [https://perma.cc/J2YJ-E3RK].

93 Robert Chesney & Danielle Keats Citron, *21st Century–Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and National Security*, 78 MD. L. REV. 882, 887–88 (2019).

94 Mary Anne Franks & Ari Ezra Waldman, *Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions*, 78 MD. L. REV. 892, 895–96 (2019).

argues that manipulated content is "a significant threat to global stability" and that "[t]hese harms are not speculative."[95] Professor Jonathan Schnader raises the concern that Internet-enabled devices such as home assistants could be compromised and that "through the audio or video obtained through Alexa, deep fake audio or video could be used as blackmail material, held ransom until a person pays money or completes a task for the blackmailer," i.e., presenting a problem with potentially direct national security implications, depending on the role of the target.[96]

A third category of manipulated information considered by the legal literature relates to fake reviews and their commercial harms. Scholars have described fake reviews as "the scourge of the reputation system"[97] and review systems as inherently manipulable.[98] Professor Eric Goldman describes a "mediated reputation system[]" as one with "third-party publisher[s] [that] gather[], organize[] and publish[] reputational information."[99] Professor Yonathan Arbel argues that in these mediated reputation systems, a mismatch exists "between the private incentives consumers have to create reputational information and its social value" and that consequently reputational information is "beset by participation, selection, and social desirability biases that systematically distort it."[100] Professor Emily Kadens explains that "[o]nline reviews . . . tend to be overwhelmingly positive or negative, with few reviews in the middle" and that this "'regression to the extreme' creates reputations that do not reflect a normal distribution of perspectives,"[101] raising questions of their trustworthiness. As explained by Professor Lori Roberts, "[S]ome businesses have gone on the offensive to preempt negative reviews to which they cannot effectively respond by incorporating non-disparagement clauses to their terms and

---

95 Nina I. Brown, *Deepfakes and the Weaponization of Disinformation*, 23 VA. J.L. & TECH. 1, 7 (2020).

96 Jonathan Schnader, *Alexa, Are You a Foreign Agent? Confronting the Risk of Foreign Intelligence Exploitation of Private Home Networks, Home Assistants, and Connectivity in the Security Clearance Process*, 25 RICH. J.L. & TECH. 3, 64 (2019).

97 Yonathan A. Arbel, *Reputation Failure: The Limits of Market Discipline in Consumer Markets*, 54 WAKE FOREST L. REV. 1239, 1295 (2019). *But see* Adi Ayal & Uri Benoliel, *Revitalizing the Case for Good Cause Statutes: The Role of Review Sites*, 19 STAN. J.L. BUS. & FIN. 331, 349 (2014).

98 Abbey Stemler, *Feedback Loop Failure: Implications for the Self-Regulation of the Sharing Economy*, 18 MINN. J.L. SCI. & TECH. 673, 676 (2017).

99 Eric Goldman, *The Regulation of Reputational Information*, *in* THE NEXT DIGITAL DECADE: ESSAYS ON THE FUTURE OF THE INTERNET 293, 294 (Berin Szoka & Adam Marcus eds., 2010).

100 *See* Arbel, *supra* note 97, at 1239.

101 Emily Kadens, *The Dark Side of Reputation*, 40 CARDOZO L. REV. 1995, 2001 (2019) (quoting Arbel, *supra* note 97, at 1265).

conditions for a purchase or service, effectively barring any negative consumer comments."[102]

Indeed, in response to the business tactic of precluding negative reviews through end user license agreements (EULAs), Congress passed the Consumer Review Fairness Act[103] in 2016.[104] In this way, Congress has begun to craft shared process-based benchmarks for conduct in a context with fake and manipulated information risks.[105] Enforcement activity under the Consumer Review Fairness Act has begun, working in concert with continued enforcement under the FTC Act.[106] For example, the FTC recently settled claims against a cosmetics company whose employees created sephora.com accounts and posted fake reviews.[107] But as substance of fakeness[108] is combined with questions of intent and quantification of harm, the analysis becomes more complex for legal purposes. In particular, legal valuation of injury and assessment of economic loss become challenging. These questions of valuation connect with the second type of manipulation—manipulation of authenticity.

---

102 Lori A. Roberts, *Brawling with the Consumer Review Site Bully*, 84 U. CIN. L. REV. 633, 637 (2016); *see also* Lucille M. Ponte, *Protecting Brand Image or Gaming the System? Consumer "Gag" Contracts in an Age of Crowdsourced Ratings and Reviews*, 7 WM. & MARY BUS. L. REV. 59, 59 (2016).

103 Consumer Review Fairness Act of 2016, 15 U.S.C. § 45b.

104 As described by the Federal Trade Commission,

> The Consumer Review Fairness Act was passed in response to reports that some businesses try to prevent people from giving honest reviews about products or services they received. Some companies put contract provisions in place, including in their online terms and conditions, that allowed them to sue or penalize consumers for posting negative reviews.

*Consumer Review Fairness Act: What Businesses Need to Know*, FTC (Feb. 2017), https://www.ftc.gov/tips-advice/business-center/guidance/consumer-review-fairness-act-what-businesses-need-know [https://perma.cc/479Y-VEER].

105 *But see* Wayne R. Barnes, *The Good, the Bad, and the Ugly of Online Reviews: The Trouble with Trolls and a Role for Contract Law After the Consumer Review Fairness Act*, 53 GA. L. REV. 549, 550 (2019).

106 *See* David Adam Friedman, *Do We Need Help Using Yelp? Regulating Advertising on Mediated Reputation Systems*, 51 U. MICH. J.L. REFORM 97, 142–43 (2017).

107 Press Release, FTC, Devumi, Owner and CEO Settle FTC Charges They Sold Fake Indicators of Social Media Influence; Cosmetics Firm Sunday Riley, CEO Settle FTC Charges that Employees Posted Fake Online Reviews at CEO's Direction (Oct. 21, 2019), https://www.ftc.gov/news-events/press-releases/2019/10/devumi-owner-ceo-settle-ftc-charges-they-sold-fake-indicators [https://perma.cc/PA95-BYE6].

108 For example, in the context of professional identity as an attorney, Professor Cassandra Robertson recommends that "states create voluntary proceedings that would allow attorneys to challenge allegedly false or misleading online reviews in a confidential setting—and would also empower disciplinary committees to impose sanctions for instances of lawyer misconduct that come to light during these proceedings." Cassandra Burke Robertson, *Online Reputation Management in Attorney Regulation*, 29 GEO. J. LEGAL ETHICS 97, 97 (2016).

### b. Manipulation of Authenticity

The New York art world harbors many tales of brilliant forgers[109] and resellers,[110] sometimes even resulting in fake art hanging on the walls of some of the most prestigious museums in the world.[111] One master of manipulating art validation was New York art dealer Ely Sakhai[112] who would acquire a minor work of a well-known master,[113] have a duplicate painted,[114] put the genuine certificate of authenticity from the original on the duplicate, and then sell the duplicate-plus-certificate on two different[115] continents.[116] The key to his success was his skillful manipulation of provenance.[117]

In perhaps surprising ways, the shortcomings of the process of art authentication and the spotting of forgeries and counterfeits are somewhat parallel to the manner in which source validation processes for information (fail to) happen on the Internet. Both the art world[118] and the Internet rely on inherently social methods of validation—

---

109 Art forgery has a proud history in the United States, dating back to the time of the Revolution. *See* ANTHONY M. AMORE, THE ART OF THE CON: THE MOST NOTORIOUS FAKES, FRAUDS, AND FORGERIES IN THE ART WORLD 1–2 (2015).

110 One such reseller was Larry Salander, a Manhattan art gallery owner. Philip Boroff, *Will Larry Salander's Fraud Victims Get Their Money Back?*, ARTNET (Apr. 18, 2014), https://news.artnet.com/market/will-larry-salanders-fraud-victims-get-their-money-back-10962 [https://perma.cc/3ML2-6FQR].

111 In one infamous case, art forger Wolfgang Beltracchi's (faked) work at one point hung in the Metropolitan Museum of Art. Joshua Hammer, *The Greatest Fake-Art Scam in History?*, VANITY FAIR (Oct. 10, 2012), https://www.vanityfair.com/culture/2012/10/wolfgang-beltracchi-helene-art-scam [https://perma.cc/6UK8-PA99].

112 *See* Dan Glaister, *Forged Gauguin Exposes Artful Dodger*, GUARDIAN (Dec. 15, 2004, 7:49 AM), https://www.theguardian.com/world/2004/dec/15/arts.usa [https://perma.cc/NGU5-AQSB].

113 The fraud spanned fifteen years and involved works by artists including Marc Chagall, Amedeo Modigliani, Pierre-Auguste Renoir, Paul Klee, and Paul Gauguin. Kelly Devine Thomas, *Update: Gallerist Goes to Prison on Art Forgery Charges*, ARTNEWS (July 19, 2005, 3:01 PM), https://www.artnews.com/art-news/news/update-gallerist-goes-to-prison-on-art-forgery-charges-1959 [https://perma.cc/G9BX-M724]; Glaister, *supra* note 112.

114 Clive Thompson, *How to Make a Fake*, N.Y. MAG. (May 20, 2004), https://nymag.com/nymetro/arts/features/9179 [https://perma.cc/K4Z8-5GN2].

115 His downfall resulted when, in 2000, both of his "Vase de Fleurs (Lilas)" by Paul Gauguin—one original, one counterfeit—were in the catalogs of London auction houses at the same time (one in Christie's catalog and one in Sotheby's). *See* Glaister, *supra* note 112. Both were listed as the lone original. *See id.* The FBI revealed that both paintings could be traced back to Sakhai. *See* Thompson, *supra* note 114.

116 The FBI estimated that his scheme grossed $3.5 million in total. *See* Glaister, *supra* note 112.

117 *See* Thompson, *supra* note 114.

118 Another infamous New York art world scandal involved the Knoedler gallery. M. H. Miller, *The Big Fake: Behind the Scenes of Knoedler Gallery's Downfall*, ARTNEWS (Apr. 25, 2016, 9:30 AM), https://www.artnews.com/art-news/artists/the-big-fake-behind-the-scenes-of-knoedler-gallerys-downfall-6179 [https://perma.cc/256J-AR8M].

methods capable of being manipulated and faked.[119] As explained by Dr. Neil Brodie,[120] an art auction can be conceptualized more as a social process than as an event; the auction house actively pairs audiences of buyers and sellers while seeking out experts to validate the objects up for sale.[121] He points out that unsavory information about the histories of objects is often not sought out or deliberately suppressed.[122]

In a similar spirit, the accurate identification of source is a question that the law regularly considers in various Internet contexts—from trademark law[123] to certain types of campaign advertisements[124] and contributions.[125] In the legal scholarship, Professor Rebecca Green has applied a counterfeiting analysis to issues of Internet campaign speech. Recalling the historical attempts to pass off forged campaign material, she argues that "post-Watergate reform addresses distribution of forged campaign material. Yet it is not clear that it would cover technology-assisted counterfeits such as deep fakes."[126] Similarly, Professor Marc Blitz has argued in favor of analyzing questions of fake news as instances of forgery, arguing that "a distinctive type of harm may arise when the falsehood is not merely in the content of the speech that is intended to deceive, but is also in its purported source or vehicle."[127] Finally, Professors Jessica Silbey and Woodrow Hartzog[128] point to insights from anthropologist Professor Graham Jones that "[t]he fake is only possible when there are normative, conventionalized, institutionalized standards of conduct and evidentiary practices that the faker can manipulate."[129] For these reasons in part, artists such as Banksy have

---

119 One example of the failure of social validation in the Internet context might be the problematic shortcoming of some certification authorities. Jake A. Berkowsky & Thaier Hayajneh, *Security Issues with Certificate Authorities*, *in* 2017 IEEE 8TH ANNUAL UBIQUITOUS COMPUTING, ELECTRONICS AND MOBILE COMMUNICATION CONFERENCE 449 (2017), https://ieeexplore.ieee.org/document/8249081 [https://perma.cc/J4LC-EEF7].

120 Brodie explains, "[W]hile auction houses can appear to be relatively passive agents in the sales process, providing a platform for bringing together buyers and sellers, in reality their business practices are more complex." Neil Brodie, *The "Art World" of the Auction Houses: The Role of Professional Experts*, 8 ARTS 1, 1 (2019), https://www.mdpi.com/2076-0752/8/2/56 [https://perma.cc/2RFG-9Y7D].

121 *Id.* at 7.

122 *Id.* at 8.

123 For a discussion of source and trademark, see, for example, Jeanne C. Fromer, *The Role of Creativity in Trademark Law*, 86 NOTRE DAME L. REV. 1885, 1887–88 (2011).

124 Rebecca Green, *Counterfeit Campaign Speech*, 70 HASTINGS L.J. 1445, 1462 (2019).

125 *Id.* at 1461.

126 *Id.* at 1470.

127 Marc Jonathan Blitz, *Lies, Line Drawing, and (Deep) Fake News*, 71 OKLA. L. REV. 59, 63–64 (2018).

128 Jessica Silbey & Woodrow Hartzog, *The Upside of Deep Fakes*, 78 MD. L. REV. 960, 966 (2019).

129 Graham M. Jones, *Deep Fakes*, *in* FAKE: ANTHROPOLOGICAL KEYWORDS 15, 21 (Jacob Copeman & Giovanni da Col eds., 2018).

begun to craft alternative social validation structures.[130] Yet, even these approaches present vulnerabilities.[131] Indeed, various failures of social validation are starkly visible among Internet fakery today.

Consider the role of celebrity endorsements and other acts of social verification on social media.[132] For example, consider the dramatic meltdown (documented in real time on social media) of the ill-fated 2017 Fyre Festival—a functionally nonexistent music festival that had been billed as "bigger than Coachella"[133] and hyped heavily by highly compensated Internet "influencers."[134] The fallout from the festival ultimately ended with the FTC sending out more than ninety letters to influencers and marketers regarding disclosure of paid endorsements[135] and with the organizer, Billy McFarland (described by the press as "the rent-a-yacht version of Frank Abagnale"),[136] being found guilty of wire fraud and sentenced to six years in prison.[137] In the aftermath, some of the influencers in question have also been sued personally[138] and reportedly were subpoenaed by the bankruptcy trustee attempting to sort through millions of unaccounted dollars that moved through Fyre Media.[139]

---

130 The street artist Banksy has set up his own validation team—Pest Control. *What Is Pest Control?*, PEST CONTROL OFF., https://pestcontroloffice.com/faq.asp [https://perma.cc/CTW8-3U34].

131 Shoshana Wodinsky, *Fake Banksy NFT Sells for $340,000 amid Suspicion the Artist's Website Was Hacked*, GIZMODO (Aug. 31, 2021, 5:20 PM), https://gizmodo.com/fake-banksy-nft-sells-for-340-000-amid-suspicion-the-a-1847593430 [https://perma.cc/HMM4-X9M8].

132 The Federal Trade Commission has issued influencer guidelines as part of an effort to mitigate the risks to consumers presented by fake endorsements driven by compensation rather than a recommendation on the merits. FTC, DISCLOSURES 101 FOR SOCIAL MEDIA INFLUENCERS (2019), https://www.ftc.gov/system/files/documents/plain-language/1001a-influencer-guide-508_1.pdf [https://perma.cc/2TML-CJCV].

133 *"The Influencers Became the Influenced"—An Industry on Fyre*, PRWEEK (Jan. 25, 2019) [hereinafter *An Industry on Fyre*], https://www.prweek.com/article/1523874/the-influencers-became-influenced-industry-fyre [https://perma.cc/M2AE-SKZ2].

134 Kendall Jenner was reportedly paid $250,000 for a single Instagram post. *Id.*

135 Press Release, FTC, FTC Staff Reminds Influencers and Brands to Clearly Disclose Relationship (Apr. 19, 2017), https://www.ftc.gov/news-events/press-releases/2017/04/ftc-staff-reminds-influencers-brands-clearly-disclose [https://perma.cc/WX7C-WAZW].

136 *See An Industry on Fyre*, *supra* note 133.

137 Doha Madani, *Fyre Festival Organizer Billy McFarland Sentenced to 6 Years on Fraud Charges*, NBC NEWS (Oct. 11, 2018, 5:13 PM), https://www.nbcnews.com/news/us-news/fyre-festival-organizer-billy-mcfarland-sentenced-6-years-fraud-charges-n919086 [https://perma.cc/6SLQ-NBKF]; Colin Moynihan, *Organizer of Failed Fyre Festival Pleads Guilty to Fraud*, N.Y. TIMES (Mar. 6, 2018), https://www.nytimes.com/2018/03/06/arts/organizer-of-failed-fyre-festival-pleads-guilty-to-fraud.html [https://perma.cc/3X7V-KVBZ].

138 *See An Industry on Fyre*, *supra* note 133.

139 Brooke Marine, *Kendall Jenner and Fyre Festival's Other Influencers Are Getting Subpoenaed*, W MAG. (Jan. 28, 2019), https://www.wmagazine.com/story/fyre-festival-bankruptcy-case-influencers-kendall-jenner-subpoena [https://perma.cc/4M5Q-7UXC].

Indeed, in situations such as that of the Fyre Festival, Internet hype and social authentication fakery start to blend into dynamics of fake personas, group pressure, and imperfect information. This blending is harnessed by the remaining MIST "Internet tarantulas" of impersonation, toxicity, and sequestration.

### 2. The Problem of Impersonation

*A fake ID works better than a Guy Fawkes mask.*[140]

The second tarantula of MIST fakery is impersonation—a fraudulent art whose history is long and storied.[141] Grifters have posed as concert promoters,[142] doctors,[143] lawyers,[144] and skilled workers of various sorts,[145] and they have regularly gone undetected. Perhaps most famously, a grifter named George C. Parker, who sometimes posed as General Grant's grandson, sold the Brooklyn Bridge, the Met, the Statue of Liberty, and Grant's Tomb to unsuspecting marks more than once.[146] Indeed, his epic grifts became legend and synonymous with the concept of smooth-talking salespeople seeking to trick unsuspecting and naïve would-be entrepreneurs with get-rich-quick fantasies.[147] His victims, unaware they had been conned, were sometimes informed of the crime as they attempted to construct tollbooths on access roads.[148]

On the Internet, we regularly encounter acts of impersonation. Twitter accounts may seem to be run by a herd of cows.[149] Instagram

---

140 V FOR VENDETTA (Warner Bros. 2005).

141 *See* Gabriella Paiella, *6 Incredible Scammers You Need to Know*, CUT (May 31, 2018), https://www.thecut.com/2018/05/6-incredible-scammer-and-grifter-stories.html [https://perma.cc/JR2R-QLR5].

142 Christian McPhate, *The Heavy Metal Grifter*, ROLLING STONE (Apr. 10, 2019), https://www.rollingstone.com/culture/culture-features/the-heavy-metal-grifter-gabe-reed-fraud-817678 [https://perma.cc/JS9G-U72F].

143 Joshua Berlinger, *Florida Teen Charged with Posing as Doctor Arrested Again*, CNN (Mar. 3, 2016, 1:39 AM), https://www.cnn.com/2016/03/03/us/malachi-love-robinson-arrest/index.html [https://perma.cc/GY65-RZMT].

144 Elie Mystal, *"Students for Trump" Grifter Pleads Guilty to Posing as a Lawyer Because of Course*, ABOVE L. (Aug. 7, 2019, 3:48 PM), https://abovethelaw.com/2019/08/students-for-trump-posed-as-lawyer-guilty [https://perma.cc/4Y5R-4TGZ].

145 Manveena Suri, *Man Arrested at Indian Airport for Impersonating Lufthansa Pilot*, CNN (Nov. 21, 2019), https://www.cnn.com/travel/article/india-fake-lufthansa-pilot-arrest/index.html [https://perma.cc/7BYF-ZLYL].

146 Gabriel Cohen, *For You, Half Price*, N.Y. TIMES (Nov. 27, 2005), https://www.nytimes.com/2005/11/27/nyregion/thecity/for-you-half-price.html [https://perma.cc/65FY-FAZD].

147 *See id.*

148 *Id.*

149 The Penn State Cows (@PSU_moo), TWITTER, https://twitter.com/PSU_moo [https://perma.cc/7PGH-3D3U].

feeds may appear to be managed by photogenic pets,[150] and pseudonymous accounts are plentiful, run by humans using privacy-preserving handles, frequently for expressive effect[151] or to protect against retaliation.[152] But, the Internet has long faced a scourge of users whose purpose is to trick others with fakery—much like grifters seeking to leverage marks. The Internet is unfortunately full of fake royalty[153] and fake humans.[154] Even the Brooklyn Bridge grift has been attempted through the Internet.[155]

While humans impersonating humans has long been a regulated area of conduct in law,[156] Internet fakery now pushes beyond the conduct of the "influencers" described in the previous Section and their manipulations of content for self-interested reasons. Today the problem is also "crime-as-a-service,"[157] sometimes designed to inflict national security harms and benefit U.S. adversaries. For example, compromised identifying information of unsuspecting humans is often purchased for purposes of online impersonation and spearphishing[158] of government employees.[159] But increasingly, new tools of technology such as machine learning are also used to generate large numbers of composite human-

150 *The Top 15 Pet Influencers Whose Instagram Followings Speak for Themselves*, MEDIAKIX, https://mediakix.com/blog/top-pet-influencers-instagram-best-popular [https://perma.cc/3VPS-23H8].

151 *See* Kate Irby, *Twitter Demands Legal Fees from Devin Nunes' Attorney in New Filing over Fake Cow's Identity*, FRESNO BEE (Feb. 6, 2020, 3:24 PM), https://www.fresnobee.com/news/local/article240046358.html (last visited Oct. 29, 2021).

152 Kurt Wagner, *Twitter Fights to Protect Anonymous Users More Often than You'd Think*, VOX (Apr. 10, 2017, 1:28 PM), https://www.vox.com/2017/4/10/15244754/twitter-lawsuit-government-anonymous-users [https://perma.cc/5AHN-29BT].

153 *Nigerian Letter or "419" Fraud*, FBI, https://www.fbi.gov/scams-and-safety/common-fraud-schemes/nigerian-letter-or-419-fraud [https://perma.cc/7FUU-27H5].

154 Peggy Anne Salz, *Bot Fraud Grows Across All Mobile Businesses and Now Threatens Apps*, FORBES (June 27, 2019, 7:16 AM), https://www.forbes.com/sites/peggyannesalz/2019/06/27/bot-fraud-grows-across-all-mobile-businesses-and-now-threatens-apps [https://perma.cc/9HC6-Q36W].

155 *See* Cohen, *supra* note 146.

156 *Identity Theft*, FBI [https://perma.cc/NQM8-5G38].

157 *See* Salz, *supra* note 154.

158 Spearphishing refers to a type of email phishing where the messages intended to trick the target are highly personalized in nature. *See, e.g.*, Microsoft 365 Team, *What Is Spear Phishing? Keep You and Your Data Safe*, MICROSOFT (Mar. 28, 2019), https://www.microsoft.com/en-us/microsoft-365/business-insights-ideas/resources/what-is-spear-phishing-how-to-keep-yourself-and-your-data-above-water [https://perma.cc/NYC2-JJ9N].

159 *What to Know About Identity Theft*, FTC (Mar. 2021), https://www.consumer.ftc.gov/articles/what-know-about-identity-theft [https://perma.cc/XHN4-AZYV]; Shannon Vavra, *Hackers Spearphished U.S. Government Agency with North Korea–Related Content Last Year*, CYBERSCOOP (Jan. 23, 2020), https://www.cyberscoop.com/government-agency-spearphishing-unit-42 [https://perma.cc/BQH2-95GQ]; David Bisson, *Former DOE Employee Pleads Guilty to Spear-Phishing Attack Against Gov't Computers*, TRIPWIRE (Feb. 4, 2016) https://www.tripwire.com/state-of-security/latest-security-news/former-doe-employee-pleads-guilty-to-spear-phishing-attack-against-govt-computers [https://perma.cc/GW97-N84B].

looking user profiles for automated impersonation.[160] Misappropriated real photos might be paired with impersonated identities.[161]

Legal scholars disagree over whether these impersonations when performed by bots are problematic and the extent to which legal intervention is warranted.[162] We find these impersonations in furtherance of fraud and national security harm problematic. Consider recent cases of foreign operatives posing as U.S. persons online, attempting to foment unrest.[163] Or consider the conduct of dating website Match.com[164] as described in the FTC enforcement action against it.[165] According to the FTC complaint, Match.com enticed users to sign up for subscriptions through the use of fake profiles expressing interest in connecting with the particular user, causing them to sign up for six months of "free" services;[166] "[t]he FTC alleges consumers often were unaware they would need to comply with additional terms to receive the free six months. . . . [and] were often billed for a six-month subscription . . . ."[167] In addition to converting their time and money, fake users such as these leave their human and corporate marks with intangible losses—feelings of annoyance, betrayal, and reputational harms, including the loss of goodwill and potential loss of intellectual

---

160 James Vincent, *ThisPersonDoesNotExist.com Uses AI to Generate Endless Fake Faces*, VERGE (Feb. 15, 2019, 7:38 AM), https://www.theverge.com/tldr/2019/2/15/18226005/ai-generated-fake-people-portraits-thispersondoesnotexist-stylegan [https://perma.cc/4FH5-RQ37].

161 *Step 1: Screen Every Profile Image Using a Reverse Online Image Search*, AUSTRALIAN CYBER SEC. CTR. (June 23, 2020), https://www.cyber.gov.au/acsc/view-all-content/guidance/step-1-screen-every-profile-image-using-reverse-online-image-search (last visited Oct. 29, 2021).

162 For a discussion of bots and fake users, see, for example, Madeline Lamo & Ryan Calo, *Regulating Bot Speech*, 66 UCLA L. REV. 988, 991 (2019) ("Crafting a narrowly tailored, enforceable law requiring bot disclosure turns out to be much harder than proponents realize, and indeed threatens to curtail an emerging form of expression.").

163 Ali Breland, *Thousands Attended Protest Organized by Russians on Facebook*, HILL (Oct. 31, 2017, 1:15 PM), https://thehill.com/policy/technology/358025-thousands-attended-protest-organized-by-russians-on-facebook [https://perma.cc/JG5M-VER8].

164 Sarah Perez, *Dating App Maker Match Sued by FTC for Fraud*, TECHCRUNCH (Sept. 26, 2019, 3:01 PM), https://techcrunch.com/2019/09/26/dating-app-maker-match-sued-by-ftc-for-fraud [https://perma.cc/9YMA-22XM].

165 Press Release, FTC, FTC Sues Owner of Online Dating Service Match.com for Using Fake Love Interest Ads to Trick Consumers into Paying for a Match.com Subscription (Sept. 25, 2019), https://www.ftc.gov/news-events/press-releases/2019/09/ftc-sues-owner-online-dating-service-matchcom-using-fake-love [https://perma.cc/BLV9-W7VM].

166 *Id.*

167 *Id.*

property.[168] In particular, trademark harms through "brandjacking"[169] might arise in this manner,[170] a new flavor of the sorts of trademark concerns that led to the passage of the Anticybersquatting Consumer Protection Act (ACPA).[171]

The challenges facing users in identifying fake information and sources also connect with the third MIST category—sequestering.

### 3. The Problem of Sequestering

> *Delivery Man: Fate whispers to the warrior.*
>
> *Ethan Hunt: A storm is coming.*
>
> *Delivery Man: And the warrior whispers back.*
>
> *Ethan Hunt: I am the storm.*[172]

The third tarantula of fakeness is information sequestering[173]—in other words, self-exacerbating information imbalances that result in both individual and group exploitation (and consequential third-party harms). Again, the law's concern over information sequestering predates the Internet; it is visible in traditional bodies of law, such as contract law[174] and tort.[175]

---

168 *See, e.g.*, Steven Melendez, *People Are Tricking Bots into Stealing Disney and Nintendo IP to Raise Awareness About Design Theft*, FAST CO. (Dec. 5, 2019), https://www.fastcompany.com/90439967/people-are-tricking-bots-into-stealing-disney-and-nintendo-ip-to-raise-awareness-about-design-theft [https://perma.cc/FRJ5-9LLK].

169 Jenny Wolfram, *Brand Impersonation on Social Media—Its Forms and Its Threats*, BRANDBASTION (Oct. 12, 2015), https://blog.brandbastion.com/brand-impersonation-on-social-media-forms-and-threats [https://perma.cc/Y748-8YZ8].

170 *What if an Instagram Account Is Using My Registered Trademark as Its Username?*, INSTAGRAM, https://help.instagram.com/101826856646059 (last visited Oct. 29, 2021).

171 For a discussion of the history of the ACPA, see, for example, Neil L. Martin, *The Anticybersquatting Consumer Protection Act: Empowering Trademark Owners, But Not the Last Word on Domain Name Disputes*, 25 J. CORP. L. 591 (2000).

172 MISSION: IMPOSSIBLE—FALLOUT (Paramount Pictures 2018).

173 The term "information sequestering" is inspired by the concept of jury sequestering, where a jury is cut off from spontaneous inputs from the outside world and presented only a carefully curated set of informational inputs in furtherance of creating a controlled environment. For a discussion of the dynamics of jury sequestering, see, for example, Marcy Strauss, *Sequestration*, 24 AM. J. CRIM. L. 63 (1996).

174 The contract law concern over a party's ability to have counsel review a contract prior to execution reflects sequestering concerns. *See, e.g.*, Eric A. Zacks, *Contract Review: Cognitive Bias, Moral Hazard, and Situational Pressure*, 9 OHIO ST. ENTREPRENEURIAL BUS. L.J. 379, 381 (2015).

175 The tort of false imprisonment involves the harms of sequestering. RESTATEMENT (SECOND) OF TORTS § 35 (1965).

a. Individual Sequestering: Informational Exploitation

The field of Internet marketing has developed progressively more sophisticated user-targeting mechanisms; social media companies "match audiences" and data brokers monitor user interactions with a high level of granularity.[176] Data aggregation and merger capabilities have advanced through technology such as facial recognition, machine learning, and various sensor-enabled data collection through the Internet of Things[177] and the Internet of Bodies.[178] Humans are now tracked both on and off the Internet, and streams of increasingly detailed data are merged in ways that are often nonobvious (to users).[179] These problems of information asymmetry have become more severe over time[180]: the content aimed at users has regularly become more tailored—curated in line with what builders of algorithms believe will elicit user engagement.

In other words, perhaps counterintuitively, from a user's perspective the Internet and related technologies have sometimes facilitated new methods of information impoverishment. They can limit rather than enhance opportunity for ready comparison of information. In the name of convenience, the algorithms of websites and apps often impose curated preferences on users, denying users visibility into the full range of options. Consider recent incidents where online credit card offers were allegedly sequestered based on gender rather than credit score,[181] and employment[182] and housing offers[183] were allegedly

176 *More Matching Capabilities with Custom Audiences*, FACEBOOK (Nov. 30, 2015), https://www.facebook.com/business/marketing-partners/partner-news/more-matching-capabilities-with-custom-audiences [https://perma.cc/QX4P-CKD5].

177 For a discussion of the legal risks of the Internet of Things, see, for example, FTC, INTERNET OF THINGS: PRIVACY & SECURITY IN A CONNECTED WORLD (2015), https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf [https://perma.cc/L67L-P5DJ].

178 For an explanation of the Internet of Bodies, see Andrea M. Matwyshyn, *The Internet of Bodies*, 61 WM. & MARY L. REV. 77, 86 (2019).

179 *Cf.* Julien Boudet, Brian Gregg, Kathryn Rathje, Eli Stein & Kai Vollhardt, *The Future of Personalization—And How to Get Ready for It*, MCKINSEY & CO. (June 18, 2019), https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/the-future-of-personalization-and-how-to-get-ready-for-it [https://perma.cc/EVM7-WHEM].

180 Uptin Saiidi, *Retailers Can Track Your Movements Inside Their Stores. Here's How*, CNBC (Mar. 7, 2019, 11:25 PM), https://www.cnbc.com/2019/03/08/how-retailers-can-track-your-movements-inside-their-stores.html [https://perma.cc/9F4W-VU2C].

181 Neil Vigdor, *Apple Card Investigated After Gender Discrimination Complaints*, N.Y. TIMES (Nov. 10, 2019), https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html [https://perma.cc/CAS7-FB7P].

182 Emily Birnbaum, *Facebook Delivers Housing, Employment Ads Based on Race, Gender Stereotypes: Study*, HILL (Apr. 4, 2019, 1:13 PM ), https://thehill.com/policy/technology/437399-facebook-delivers-housing-employment-ads-based-on-race-and-gender [https://perma.cc/6KMW-BSG6].

183 *Id.*

sequestered (directly or indirectly) based on race.[184] Thus, while the Internet appears to offer a glut of information, tracking technologies distort available information by crafting an opportunity-limited, sequestered Internet experience for selected users. Although it now implicates twenty-first century technologies, the problem of artificial information sequestration itself is not a new concern for law.[185] Indeed, contract law has long considered the risks of sequestration as a method of deception during the process of contract formation, potentially rendering formation invalid.[186]

For example, in the case of election information sequestering, former Cambridge Analytica employees explain that the same election advertisement can be morphed into thousands of variants in order to target particular individuals in ways believed to be more resonant.[187] These dynamics became evident when, after congressional pressure, Facebook launched a campaign advertising archive,[188] which permitted users to review versions of advertisements run on Facebook by various candidates for office.[189] This archive coupled with disclosures from whistleblowers reveal a high level of personalization and A/B testing in honing sometimes seemingly contradictory political messages.[190] This extent of user targeting in political messaging not only raises concerns with respect to information sequestering, it also highlights the existence of a toolbox of techniques that are likely effective at exploiting users' lack of skill in detecting social engineering and potential fakery. These tools also, in turn, feed the creation of increased social sequestration and polarization. In other words, what is lost in this process is the

184 *Id.*

185 For example, in contract law contexts, courts analyze the doctrine of duress to include inquiries into whether a party with superior bargaining power generated a false sense of urgency in the transaction or created circumstances designed to prevent the other party from seeking third party counsel and outside information. *See, e.g.*, Odorizzi v. Bloomfield Sch. Dist., 54 Cal. Rptr. 533 (Cal. Ct. App. 1966). In criminal law contexts, the act of assuming custodial control over someone in a manner that limits access to outside information and parties has raised concern with courts in certain circumstances. *See, e.g.*, Rachele Norfolk, *Solving the Depraved Heart Murder Problem in Maryland: A Suggestion for Successful Prosecution of Police Officers*, 46 U. BALT. L. REV. 547, 558–60 (2017) (discussing depraved heart murder in police custody and the role of sequestration and pleas for medical assistance).

186 *See* Zacks, *supra* note 174, at 386.

187 John Naughton, *Cambridge Analytica: Mindf*ck by Christopher Wylie; Targeted by Brittany Kaiser—Reviews*, GUARDIAN (Oct. 29, 2019, 2:00 AM), https://www.theguardian.com/books/2019/oct/29/mindfck-christopher-wylie-targeted-brittany-kaiser-cambridge-analytica-review [https://perma.cc/5YP4-J8Y6].

188 Josh Constine, *Facebook Launches Searchable Transparency Library of All Active Ads*, TECHCRUNCH (Mar. 28, 2019, 7:00 PM), https://techcrunch.com/2019/03/28/facebook-ads-library [https://perma.cc/TF5Z-LW3W]; *see also* Mike Isaac, *Why Everyone Is Angry at Facebook over Its Political Ads Policy*, N.Y. TIMES (Nov. 22, 2019), https://www.nytimes.com/2019/11/22/technology/campaigns-pressure-facebook-political-ads.html [https://perma.cc/M38Z-UHHW].

189 *Ad Library*, FACEBOOK, https://www.facebook.com/ads/library (last visited Oct. 29, 2021).

190 *See* Naughton, *supra* note 187.

creation of a shared user experience in both the offline and online world—a common frame of reference.

b. Social Sequestering: Alternative Belief Systems

The second category of sequestering involves dynamics of group isolation and the collective version of what Professor Cass Sunstein has referred to as Daily Me[191] echo chambers—self-reinforcing groups who create alternative belief systems driven by factually unsupported beliefs. As these information silos develop progressively more elaborate belief systems, their members may disconnect from people outside the group; because of their reliance on fake information pushed through tech-enabled mechanisms of social sequestering, they experience (and cause) various categories of potentially legally problematic content and conduct. In particular, their engagement on the Internet increasingly involves recognizable memes and tropes that leverage social sequestration to further reinforce a crafted group identity. For example, members of Internet conspiracy groups often describe feelings of alienation from family and friends.[192] Indeed, these dynamics of social sequestering bear some resemblance to former participant descriptions of informational dynamics of cults[193] on the one hand and pyramid/multi-level marketing (MLM) schemes[194] on the other. Again, perhaps much like cults and pyramid/MLM schemes, the harms suffered by these groups include increased susceptibility to technology-assisted financial exploitation. For example, some media personalities who arguably leverage social sequestration dynamics have sold fake products online to their audiences, triggering warnings from federal agencies.[195]

191 CASS SUNSTEIN, REPUBLIC.COM 7–12 (2001).

192 Travis M. Andrews, *QAnon Is Tearing Families Apart*, WASH. POST (Oct. 12, 2020, 6:00 AM), https://www.washingtonpost.com/technology/2020/09/14/qanon-families-support-group [https://perma.cc/R34F-NZTU].

193 *See, e.g.*, Laura B. Brown, *He Who Controls the Mind Controls the Body: False Imprisonment, Religious Cults, and the Destruction of Volitional Capacity*, 25 VAL. U. L. REV. 407, 407–09 (1991).

194 *See, e.g.*, Sergio Pareja, *Sales Gone Wild: Will the FTC's Business Opportunity Rule Put an End to Pyramid Marketing Schemes?*, 39 MCGEORGE L. REV. 83, 84 (2008).

195 Warning Letter from Donald D. Ashley, Dir., Off. of Compliance, Ctr. for Drug Evaluation & Rsch., FDA. & Richard A. Quaresima, Acting Assoc. Dir., Div. of Advert. Pracs., FTC, to Alexander E. Jones, Free Speech Sys. LLC (Apr. 9, 2020), https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/free-speech-systems-llc-dba-infowarscom-605802-04092020 [https://perma.cc/UQA7-DE6Z].

### 4. The Problem of Toxicity

> *Martha grew flowers. . . . They were beautiful flowers, and their scent entranced. But, however beautiful, these flowers were also poisonous.*[196]

The fourth tarantula of Internet fakery is toxicity.[197] Toxicity on the Internet is not new; it has caused concern for users and scholars alike since the Internet's earliest days. For example, as early as 1993, authors described virtual "rape" that disrupted Internet communities.[198] From the racist memes of 8chan[199] to the group harassment of Gamergate[200] to posts of nonconsensual pornography[201] and child exploitation content,[202] the dark side of the Internet causes even the most zealous First Amendment defenders to recognize the negative consequences of unbridled technology-assisted toxicity.[203] Toxic environments are not new to law and are considered literally in environmental law, but also both in physical[204] and psychological[205] terms in employment and labor law.

---

196 LAWRENCE LESSIG, CODE: VERSION 2.0, at 10 (2006).

197 In this case, toxicity refers to the quality of an environment where users perceive themselves to have disincentives to engage due to emergent, negative experiences whose harsh or harmful quality is severe enough to potentially outweigh the users' perception of the benefits of engagement.

198 Julian Dibbell, *A Rape in Cyberspace*, VILL. VOICE (Oct. 18, 2005), https://www.villagevoice.com/2005/10/18/a-rape-in-cyberspace [https://perma.cc/ZGC5-R6MR].

199 *8chan/8kun*, KNOW YOUR MEME, https://knowyourmeme.com/memes/sites/8chan-8kun [https://perma.cc/XRK3-62HU].

200 Caitlin Dewey, *The Only Guide to Gamergate You Will Ever Need to Read*, WASH. POST (Oct. 14, 2014, 5:23 PM), https://www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read [https://perma.cc/6F2Q-B53Y].

201 *What to Do if You're the Target of Revenge Porn*, FTC (May 2021), https://www.consumer.ftc.gov/blog/2018/01/what-do-if-youre-target-revenge-porn [https://perma.cc/HT2A-8FE4].

202 Michael H. Keller & Gabriel J.X. Dance, *The Internet Is Overrun with Images of Child Sexual Abuse. What Went Wrong?*, N.Y. TIMES (Sept. 29, 2019), https://www.nytimes.com/interactive/2019/09/28/us/child-sex-abuse.html [https://perma.cc/Q79P-YKXA].

203 *See* Emily Bazelon, *Lori Drew Is a Meanie*, SLATE (Dec. 3, 2008, 6:33 PM), https://slate.com/news-and-politics/2008/12/lori-drew-s-conviction-in-the-myspace-suicide-case.html [https://perma.cc/L5XF-JNWT].

204 For a discussion of physically toxic work environments, see, for example, Hannah Arterian Furnish, *Beyond Protection: Relevant Difference and Equality in the Toxic Work Environment*, 21 U.C. DAVIS L. REV. 1, 3 (1987).

205 For a discussion of emotionally toxic environments, see, for example, Vicki Schultz, *Reconceptualizing Sexual Harassment*, 107 YALE L.J. 1683, 1687 (1998).

Yet, the fit between existing law and recourse for technology toxicity is not always ideal.[206] Consider the tragic case of Megan Meier, a Missouri teen who committed suicide after orchestrated bullying and verbal abuse by a group of her classmates and an adult, Lori Drew.[207] While the Missouri court struggled to apply Missouri law at the time, California prosecutors relied on a novel theory under the Computer Fraud and Abuse Act—the idea that a mere breach of contract can present the basis for a criminal charge of computer intrusion—to bring charges against the defendant.[208] Although a jury ultimately convicted the defendant in *United States v. Drew*,[209] the judgment was set aside through a judgment notwithstanding the verdict,[210] and scholars have critiqued the theory of the case crafted by prosecutors.[211] Regardless of whether scholars agree with the court's analysis, however, the facts of *United States v. Drew* illustrate how toxic Internet exchanges can seep into the consciousness of Internet users, with physical harms sometimes following.

Today, Internet toxicity concerns are perhaps most readily visible in the context of the ongoing debate over whether to amend Section 230 of Part I of Subchapter II of Chapter 5 of Title 47, commonly referenced as "Communications Decency Act Section 230" or "CDA 230."[212] Passed in 1996, CDA 230 is often credited with stimulating the rapid growth of Internet business and content of the last two decades.[213] However, the trade-off for the buffers of liability protection of CDA

---

206 In response to the *Drew* case, Missouri passed a new law addressing cyberbullying. Associated Press, *Missouri: Cyberbullying Law Is Signed*, N.Y. TIMES (July 1, 2008), https://www.nytimes.com/2008/07/01/us/01brfs-CYBERBULLYIN_BRF.html [https://perma.cc/WYS6-9TQX].

207 Leocadie Welling, United States v. Drew*: District Court Judge Rules Evidence of Suicide Admissible in Lori Drew MySpace Case*, HARV. J.L. & TECH. DIG. (2008), http://jolt.law.harvard.edu/digest/united-states-v-drew [https://perma.cc/B6M3-3WUR].

208 United States v. Drew, 259 F.R.D. 449, 451 (C.D. Cal. 2009). For a discussion of the theory of the case, see Andrea M. Matwyshyn & Stephanie K. Pell, *Broken*, 32 HARV. J.L. & TECH. 479, 485–87 (2019).

209 Jennifer Steinhauer, *Verdict in MySpace Suicide Case*, N.Y. TIMES (Nov. 26, 2008), https://www.nytimes.com/2008/11/27/us/27myspace.html [https://perma.cc/W2RH-N4R8].

210 Bobbie Johnson, *Judge Overturns Guilty Verdict in MySpace Suicide Case*, GUARDIAN (July 2, 2009, 3:50 PM), https://www.theguardian.com/technology/blog/2009/jul/02/lori-drew-myspace-acquitted [https://perma.cc/JUC5-C4CL].

211 *See, e.g.*, Brief of Professor Orin S. Kerr as Amicus Curiae in Support of Petitioner at 18, Van Buren v. United States, 141 S. Ct. 1648 (2021) (No. 19-783).

212 47 U.S.C. § 230.

213 Joshua A. Geltzer, *The President and Congress Are Thinking of Changing This Important Internet Law*, SLATE (Feb. 25, 2019, 3:40 PM), https://slate.com/technology/2019/02/cda-section-230-trump-congress.html [https://perma.cc/X2BP-VHC2]; Adi Robertson, *Why the Internet's Most Important Law Exists and How People Are Still Getting It Wrong*, VERGE (June 21, 2019, 1:02 PM), https://www.theverge.com/2019/6/21/18700605/section-230-Internet-law-twenty-six-words-that-created-the-Internet-jeff-kosseff-interview [https://perma.cc/HYQ4-T6WH].

230[214] has been in part the proliferation of fake content and conduct that is the subject of this Article.[215] While digital copyright issues were provided additional consideration separately in the Digital Millennium Copyright Act (DMCA),[216] the language of CDA 230 does not include the DMCA's updating mechanisms through the Library of Congress. As such, robust calls for[217] (and against)[218] amendments to CDA 230 have become louder as the toxicity of Internet fakeness has increased.

Although these problems of MIST map to traditional legal corollaries, they are exacerbated by two dynamics relating to data aggregation and leveraging capabilities that are central to (the current version of) the Internet—the arrival of the *Internet "long con"* and the risk of the rise of what we term the *"PSYOP industrial complex."*

### B. *Internet Con(tent and Conduct)*

Before the exploits of con artist Frank Abagnale, law enforcement faced the grifts of Ferdinand Demara, also known as the Great Impostor.[219] Demara grifted through life since the age of sixteen under numerous identities.[220] He joined the Navy,[221] faked his suicide,[222] became a psychologist,[223] performed surgeries as a doctor,[224] taught as a

---

214 CDA 230 provides, in relevant part:

> No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider. . . .
>
> No provider or user of an interactive computer service shall be held liable on account of . . . any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.

47 U.S.C. § 230(c)(1)–(2)(A).

215 The specifics of fakeness and intermediation are outside the scope of this Article. They raise additional concerns that will be considered in a subsequent article.

216 For a discussion of the DMCA Section 1201 Rulemaking process, see *Rulemaking Proceedings Under Section 1201 of Title 17*, U.S. COPYRIGHT OFF., https://www.copyright.gov/1201 [https://perma.cc/4YCV-4W9V].

217 Frank Ready, *Timing Is Ripe for Section 230 Amendments—but the "How" Is Missing*, LAW.COM (Oct. 24, 2019, 11:00 AM), https://www.law.com/legaltechnews/2019/10/24/timing-is-ripe-for-section-230-amendments-but-the-how-is-missing [https://perma.cc/ED7R-HRLR].

218 Elliot Harmon, *Changing Section 230 Would Strengthen the Biggest Tech Companies*, N.Y. TIMES (Oct. 16, 2019), https://www.nytimes.com/2019/10/16/opinion/section-230-freedom-speech.html [https://perma.cc/7QSX-UBNL].

219 ROBERT CRICHTON, THE GREAT IMPOSTER (1959).

220 *Id.* at 5–6.

221 *Id.* at 57.

222 *Id.* at 65.

223 *Id.*

224 *Id.* at 148.

college professor,[225] founded a college,[226] and even studied law.[227] Later, he explained that the secret to his success was driven by two beliefs: "One was that in any organization there is always a lot of loose, unused power lying about which can be picked up without alienating anyone. The second rule is, if you want power and want to expand, never encroach on anyone else's domain; open up new ones."[228]

Demara's two principles for "[e]xpanding into the power vacuum"[229] may explain a portion of the dynamics driving technology fakery today. The first dynamic of harnessing "loose" power can be seen in the merger of the state of the art of offline social engineering with the state of the art of technology-assisted data aggregation—a merger that targets "marks" more effectively. We might term this dynamic the arrival of the Internet "long con." The second dynamic, opening new domains, can be seen in market movement toward a (problematic) merger between private sector information-targeting capabilities and the techniques of psychological operations from military contexts—what we term the "PSYOP industrial complex."

### 1. The Internet Long Con

> *The crook is always attracted to regions of sudden prosperity and quick expansion. There he finds loose and easy money.*[230]
>
> —*Edward H. Smith*

In 1923, con artist Edward H. Smith recounted the theory and folk history of classical U.S. con artistry in his work, *Confessions of a Confidence Man: A Handbook for Suckers*.[231] Smith explains that "[c]onfidence is a business, and, like all business, changes and conforms to conditions. In fact, con takes rise from the conditions of life about it and adapts itself as does social life."[232] Thus, it should perhaps be no surprise that as the Internet era became the new gold rush, grifting and scamming evolved to include it. For example, traditional romance scams once again caught the attention of the Federal Trade Commission as they successfully shifted online: instead of a con artist actively

---

225 *Id.* at 93.

226 *Id.* at 92.

227 *Id.* at 109.

228 *Id.* at 102–03.

229 *Id.* at 103.

230 EDWARD H. SMITH, CONFESSIONS OF A CONFIDENCE MAN: A HANDBOOK FOR SUCKERS 7 (1922).

231 *Id.*

232 *Id.* at 9.

cultivating a relationship in physical space,[233] a version of the story arose with an Internet "sweetheart" (or catfish)[234] living in another country without the funds to leave.[235]

Cons divide into two categories—short and long/big.[236] As explained by Luc Sante, the short con is a one-shot interaction in which the mark is tricked out of the money on hand, whereas the long con involves multiple interactions; in the classic long con, "a form of theater . . . staged . . . for an audience of one, who is moreover enlisted as part of the cast," the victim is sent home to get more money to lose.[237] The cons described in Smith's book unfold across days or weeks, often involving multiple scenes and actors.[238] Indeed, Smith describes a well-conceived confidence game as including at least five separate steps prior to the ultimate payoff: foundation work,[239] the approach,[240] the build-up,[241] the payoff or convincer,[242] the "hurrah,"[243] and sometimes also the in-and-in[244] and the corroboration.[245]

For the purposes of a con artist's foundation work in particular, ease of access to information about the mark becomes another point of the mark's vulnerability. As the Federal Bureau of Investigation has explained, the Internet has provided new avenues for con artists; information on social media sites and other Internet sources means that the con artist can craft a plan for generating a false sense of familiarity

---

233 *Man "Posed as MI6 Spy to Con Woman Out of £850,000,"* TELEGRAPH (Oct. 19, 2016, 3:34 PM), https://www.telegraph.co.uk/news/2016/10/19/man-who-posed-as-mi6-spy-hunted-over-claims-he-conned-divorcee-o [https://perma.cc/FZJ2-EAAN].

234 Molly McHugh, *The Life of an Internet Catfish Is Rough These Days*, WIRED (July 7, 2015, 8:00 AM), https://www.wired.com/2015/07/catfish-2 [https://perma.cc/A79Z-TFCB].

235 Emma Fletcher, *Romance Scams Rank Number One on Total Reported Losses*, FTC (Feb. 12, 2019, 9:23 AM), https://www.ftc.gov/news-events/blogs/data-spotlight/2019/02/romance-scams-rank-number-one-total-reported-losses [https://perma.cc/7ZKV-YLNN].

236 Luc Sante, *Introduction* to DAVID W. MAURER, THE BIG CON: THE STORY OF THE CONFIDENCE MAN, at vii, x (Anchor Books 1999) (1940).

237 *Id.* at xi.

238 *See* SMITH, *supra* note 230.

239 Foundation work refers to preparations; in particular, studying the background knowledge needed for playing the role and for conning the mark. *Id.* at 35–36.

240 The approach refers to the method of contacting the target of the con. *Id.* at 36.

241 The build-up involves offering the mark the opportunity to profit from a scheme, encouraging greed over rational judgment. *Id.*

242 The payoff or convincer is a payout as proof of the scheme's purported effectiveness—either in money, or faked. *Id.*

243 The "hurrah" is a sudden manufactured crisis or change of events that forces the mark to act or to make a decision immediately. *Id.*

244 The in-and-in is a step where a confederate puts some money into the same scheme as the mark, to add an appearance of legitimacy. *Id.*

245 The corroboration step involves a confirmation of claims made by the con man by someone who appears to be an uninvolved third party, but in reality may not be so. *Id.* at 37.

with the mark through background research.[246] Data resale companies offer "background check" information on any person[247] of a con artist's choosing, i.e., extensive reconnaissance and background reading, full of clickstream data, purchases, places of former employment, past addresses, networks of friends, and other information, all of which may prove potentially useful in a con. As explained in an interview with one expert, "Today the con artist's job is easier than ever because much of the work—gathering information about a potential mark—is done for him by people who voluntarily check in wherever they go."[248] In other words, self-disclosed Internet information offers valuable insights into a potential mark's preferences and the kind of confidence games that might succeed.[249] Additionally, the Internet potentially allows some aspects of the con to be automated as well as personalized.[250] Thus, cons become less labor-intensive to run, potentially allowing the simultaneous manipulation of a larger number of marks than may be practical in classic face-to-face cons.

Smith also explains that one reason for the success of cons is that they "play[] an invariable chord in the human make-up—good old earthy greed."[251] He continues: "There are other reasons why con is perennial. It has taken advantage from the beginning of the public foibles, of what is now termed mass psychology."[252] This observation connects us to two other sets of social engineers—marketers and propagandists/PSYOP professionals—and their uses of increasingly granular information-targeting capabilities for goal-oriented exploitation of human psychology.

---

246 *Romance Scams*, FBI, https://www.fbi.gov/scams-and-safety/common-scams-and-crimes/romance-scams [https://perma.cc/J2U2-7HKS] ("Be careful what you post and make public online. Scammers can use details shared on social media and dating sites to better understand and target you.").

247 *Start Your People Search Today!*, INTELIUS, https://www.intelius.com/people-search (last visited Nov. 19, 2021).

248 Angela Chen, *The Art of the Con: Maria Konnikova on Scams, Grifters and Being Easily Duped*, GUARDIAN (Jan. 15, 2016, 11:24 AM), https://www.theguardian.com/books/2016/jan/15/maria-konnikova-interview-new-book-the-confidence-game-review-scams [https://perma.cc/7U4H-NN9U].

249 *Id.*

250 *See* Perez, *supra* note 164.

251 SMITH, *supra* note 230, at 9–10.

252 *Id.* at 10.

### 2. The PSYOP Industrial Complex

> *Yet, in holding scientific research and discovery in respect, as we should, we must also be alert to the equal and opposite danger that public policy could itself become the captive of a scientific-technological elite.*
>
> *—President Dwight D. Eisenhower*[253]

During World War II, the German Army blanketed its own forces with leaflets, attempting to boost morale among its troops.[254] Called Skorpion, the leaflets discussed new "super weapons" and the hope of German victory.[255] However, Allied psychological warfare personnel obtained copies of Skorpion and prepared their own fake version of the leaflets—a version with an Allied slant to the information but was otherwise identical.[256] The desired effect succeeded: after the Allies airdropped millions of these fake leaflets on German troops in the field, the true Skorpion leaflets were soon discontinued.[257] These are the fakery techniques of the military field of psychological operations or PSYOP.[258]

As explained by the Army Field Manual in *Psychological Operations Tactics, Techniques, and Procedures*,

> PSYOP are planned operations that convey selected information and indicators to foreign target audiences (TAs) to influence their emotions, motives, objective reasoning, and ultimately, the behavior of foreign governments, organizations, groups, and individuals. The purpose of all

253 President Dwight D. Eisenhower, Farewell Address (Jan. 17, 1961), https://www.ourdocuments.gov/doc.php?flash=false&doc=90&page=transcript [https://perma.cc/6DKC-CQ52].

254 HEADQUARTERS, DEP'T OF THE ARMY, PSYCHOLOGICAL OPERATIONS TACTICS, TECHNIQUES, AND PROCEDURES 11-24 (2003), https://fas.org/irp/doddir/army/fm3-05-301.pdf [https://perma.cc/LP34-VNAZ].

255 *Id.*

256 These revised leaflets included content such as instructions to German troops to "shoot their officers if they did not display sufficient 'National Socialist' zeal." *Id.*

257 *Id.*

258 *Id.* at 1-1.

> PSYOP[259] is to create in neutral, friendly, or hostile foreign groups the emotions, attitudes, or desired behavior . . . .[260]

The Army further explains that PSYOP techniques are ever-changing[261] and "[p]roven in combat and peacetime."[262] They are "one of the oldest weapons in the arsenal . . . , as well as an important force protector, combat multiplier, and nonlethal weapons system."[263] Indeed, the story of the Skorpion leaflets reminds us, strategic fake information injection has long been a staple tool of military operations. Fakery is not new or specific to the Internet.

Legal scholars usually frame discussions of fake Internet information and "disinformation" under varying (nonmilitary) definitions and without clear engagement with this complex militarized history.[264] The history of the field of PSYOP reveals the import of filling in this gap in existing scholarly work.[265] Almost all existing legal scholarship analyzes Internet fakery in a compartmentalized way—as something either political on the one hand or economic on the other,[266] but this segmentation misframes the problem. The history of PSYOP reveals a clear *absence* of sectoral segmentation. PSYOP has frequently reached into the private sector for the state of the art of psychology and advertising industry knowledge.[267] In other words, the *interwoven*

---

259 *Id.* at 1-1. PSYOP are offensive information operations that serve to advise commanders on future action toward successful mission completion, to "influence foreign populations by expressing information subjectively [in ways that] influence attitudes and behavior," to facilitate military operations, to provide public information to foreign populations, to "serve as the supported commander's voice to foreign populations to convey intent and establish credibility," and to "counter enemy propaganda, misinformation, disinformation, and opposing information" and other related goals. *Id.* at 1-2 to 1-3.

260 *Id.* at 1-1.

261 *Id.*

262 *Id.* at 1-2.

263 *Id.*

264 As explained by one scholar: "The recent usage of the term [fake news] . . . shows that the term has been used in different meanings in the past by scholars. . . . [and] has no coherent meaning. . . . Disinformation, misinformation, and propaganda all have somehow similar meanings as 'fake news.'" Andrei Richter, *Fake News and Freedom of the Media*, 8 J. INT'L MEDIA & ENT. L. 1, 8–9 (2018).

265 *See supra* notes 27–36 and accompanying text.

266 These intuitions may arise from the traditional First Amendment approach to speech which historically differentiated between political and commercial speech. However, this distinction grows exceedingly tenuous under current Supreme Court case law. For a discussion of the merger of commercial and policy speech, see, for example, Martin H. Redish, *Commercial Speech, First Amendment Intuitionism and the Twilight Zone of Viewpoint Discrimination* (Northwestern Univ. Sch. of L., Fac. Working Paper No. 155, 2008), https://scholarlycommons.law.northwestern.edu/facultyworkingpapers/155 [https://perma.cc/JUQ2-4SUT].

267 *See, e.g.*, HEADQUARTERS, *supra* note 254, at 11–23 (making reference to Bill Gates's public relations communication techniques in refuting expert conclusions presented during the Microsoft antitrust trial).

*nature of private sector and public sector in PSYOP* and psychographic propaganda efforts has been a historical reality for at least a century in both the United States and elsewhere. Framed in the modern language of information security, the dynamic reflects an early form of what one of us has elsewhere described as the problem of "reciprocal security vulnerability."[268] The information security dynamics of the private sector influence the public sector and vice versa. To wit, a comprehensive legal paradigm for Internet fakery should explicitly *recognize and consider both economic and political Internet fakery dynamics within part of a single construct.*

An early example of the merger of various private sector psychographic techniques with psychological operations is visible in the writing of Edward L. Bernays, a founder of modern advertising messaging[269] (and perhaps the original user of media "influencers").[270] Bernays served during World War I as an integral part of the U.S. Committee on Public Information (CPI) to message the war effort.[271] Indeed, he coined the term "engineering of consent,"[272] and his seminal

---

268 Andrea M. Matwyshyn, *CYBER!*, 2017 BYU L. REV. 1109, 1113 (referring to "reciprocal security vulnerability" as the problem that "the practical reality that the information security of the private and public sector are inextricably interwoven").

269 *Edward Bernays, "Father of Public Relations" and Leader in Opinion Making, Dies at 103*, N.Y. TIMES (Mar. 10, 1995), https://archive.nytimes.com/www.nytimes.com/books/98/08/16/specials/bernays-obit.html [https://perma.cc/LY7D-AJNV]; *see also, e.g.*, Jeremy Geltzer, *Fake News & Film: How Alternative Facts Influence the National Discourse*, 47 SW. L. REV. 297, 305, 309–10 (2018) ("Bernays found that two of the most useful tools in the art of coercion were fear and fashion. . . . [H]is book, *Crystallizing Public Opinion*, occupied a privileged position in the library of Nazi Propaganda Minister Joseph Goebbels."); Christyne J. Vachon, *Crocodile Tears: How Businesses Use Animal Testing Labeling as Propaganda to Increase Profit*, 14 J. ANIMAL & NAT. RES. L. 179, 191 (2018) (crediting Bernays with recognizing that "[e]verything about public relations is about selling an idea"); William C. Tucker, *The Big Lie: Is Climate Change Denial a Crime Against Humanity?*, 7 INTERDISC. J. HUM. RTS. L. 91, 100 (2013) (pointing to Bernays's statement that the public relations professional "[s]hould be candid in his dealings . . . his business is not to fool or hoodwink the public" (ellipses in original)); Tamara R. Piety, *Free Advertising: The Case for Public Relations as Commercial Speech*, 10 LEWIS & CLARK L. REV. 367, 400 (2006) ("[I]nvisible government was necessary according to Bernays because, in order for the theory of competitive markets to work in the face of the reality of a surfeit of information and a tendency for people therefore to follow tastemakers, those minorities who actually acted as the governors needed a device to mold the mind of the masses [so] that they will throw their newly gained strength in the desired direction." (internal quotations omitted) (alteration in original)).

270 When running a campaign to promote consumption of bacon, Bernays enlisted doctors to recommend a "hearty" breakfast. Lisa Held, *Psychoanalysis Shapes Consumer Culture*, AM. PSYCH. ASS'N, Dec. 2009, at 32, https://www.apa.org/monitor/2009/12/consumer [https://perma.cc/LYM5-EKUS].

271 Christopher B. Daly, *How Woodrow Wilson's Propaganda Machine Changed American Journalism*, SMITHSONIAN MAG. (Apr. 28, 2017), https://www.smithsonianmag.com/history/how-woodrow-wilsons-propaganda-machine-changed-american-journalism-180963082 [https://perma.cc/BR48-9B2S].

272 Edward L. Bernays, *The Engineering of Consent*, 250 ANNALS AM. ACAD. POL. & SOC. SCI. 113, 114 (1947).

work, *Propaganda*,[273] blends insights from Freudian group psychology with behaviorist principles.[274] In the book's opening sentence, channeling his hybrid experiences in both government and advertising settings, Bernays states that "[t]he conscious and intelligent manipulation of the organized habits and opinions of the masses is an important element in democratic society."[275] He highlights the role that messaging reinforcement and repetition play in maximizing efficacy,[276] stating that "[p]ropaganda is the executive arm of the invisible government."[277] But, today's problems of Internet fakery add a new wrinkle. They are rooted not only in this "pull" from the private sector into military PSYOP; they also evidence the arrival of the inverse relationship from military settings into commercial ones.[278] A new industry of disinformation-for-hire or "dark PR"[279] has arisen, used not only by governments[280] but also by companies, celebrities, and others.[281] The clients of these dark PR firms are knowingly hiring ex-PSYOP operatives to engage in targeted disinformation operations[282] on their

---

273 EDWARD L. BERNAYS, PROPAGANDA (1928).

274 Held, *supra* note 270.

275 BERNAYS, *supra* note 273, at 9.

276 Bernays's insights on the efficacy of messaging repetition might be said to conceptually connect with what an economist might call a "sunk cost" of prior subscription, potentially leading to a (misplaced) unwillingness to reconsider beliefs. Social psychologists and communication scholars might add the multiplying impact of reinforcement if the language used to reinforce the message connects with the target's group identity. *See, e.g.*, Henri Tajfel & John Turner, *"An Integrative Theory of Intergroup Conflict," in* THE SOCIAL PSYCHOLOGY OF INTERGROUP RELATIONS 33, 33 (William G. Austin & Stephen Worchel eds., 1979); Daniel S. Lane, Stewart M. Coles & Muniba Saleem, *Solidarity Effects in Social Movement Messaging: How Cueing Dominant Group Identity Can Increase Movement Support*, 45 HUM. COMMC'N RSCH. 1, 1 (2019).

277 BERNAYS, *supra* note 273, at 20.

278 To give one example, the website of the business services company Black Cube describes its team as "[h]ighly experienced and trained in Israel's elite military and governmental intelligence units," who offer "several unique methods, especially in the social engineering field" to solve complex business and litigation issues for Black Cube's clients, including "extract[ing] valuable information from limited access sources" in virtual environments. BLACK CUBE, http://www.blackcube.com [https://perma.cc/8X67-CHZQ].

279 This Article refers to the fledgling industry as "dark PR," merging the concept of the "dark arts" of computer intrusion/technology surveillance and PSYOP with marketing; marketing professionals more often refer to this industry as "black PR," presumably linguistically merging the concepts of black hat hacking with traditional PR. Thus, the term dark PR reframes the concept slightly more broadly.

280 USENIX Enigma Conference, *USENIX Enigma 2020—Disinformation (Panel)*, YOUTUBE, at 44:00 (Mar. 10, 2020), https://youtu.be/4iXFxT_4cO0?t=2640 (last visited Nov. 20, 2021).

281 *Id.* at 30:30.

282 Further, private sector entities, including such dark PR firms, are now sometimes aided by state-of-the-art offensive technology surveillance tools originally created for governments. For example, Facebook has recently sued NSO Group, a maker of government surveillance software used by governments, in connection with the alleged exploitation of an audio vulnerability. *See, e.g.*, Dana Priest & Elizabeth Dwoskin, *Chief of WhatsApp, Which Sued NSO over Alleged Hacking*

behalf.[283] But, in addition to the dark PR industry and a broader revolving door of PSYOP-trained personnel between military and civilian contexts,[284] there are potentially broader parallel dynamics between military PSYOP techniques and the common audience-targeting techniques of today's Internet advertising infrastructure in general.[285] Thus, today's Internet fakery includes a new reciprocal pull where government PSYOP techniques appear to be leaking into the private sector.[286] In brief, *Internet fakery today potentially arises in part from PSYOP techniques beginning to creep into civilian contexts*.

Marketing professionals themselves are ethically troubled by this creep. In the words of one such communications professional, the current Internet operations of dark PR firms and related private sector dynamics have constructed an "ecosystem that is just so ripe for professional lying."[287] Further, marketing executives explain, the new norms of extreme microtargeting exploitation conflict with the traditional ethical boundaries of pre-Internet marketing.[288] The transgression of this traditional ethical boundary when analyzed in

---

*of Its Product, Disputes Firm's Denials on Scope of, Involvement in Spyware Operations*, WASH. POST (July 24, 2021, 11:27 AM), https://www.washingtonpost.com/investigations/2021/07/24/whatsapp-pegasus-spyware [https://perma.cc/LCW7-S6YK] ("In court documents, NSO has argued that it should be granted 'sovereign immunity' because its clients are vetted government customers, and legal doctrine holds that governments cannot be sued for performing their legitimate functions."). The same NSO Group tools are sometimes also sought out by private sector entities. Devin Coldewey, *Before Suing NSO Group, Facebook Allegedly Sought Their Software to Better Spy on Users*, TECHCRUNCH (Apr. 3, 2020, 7:37 PM), https://techcrunch.com/2020/04/03/before-suing-nso-group-facebook-allegedly-sought-their-software-to-better-spy-on-users [https://perma.cc/R9SU-V5HY] ("But before complaining about the company's methods, Facebook seems to have wanted to use them for its own purposes, according to testimony from NSO founder Shalev Hulio.").

283 Adam Entous & Ronan Farrow, *Private Mossad for Hire*, NEW YORKER (Feb. 11, 2019), https://www.newyorker.com/magazine/2019/02/18/private-mossad-for-hire [https://perma.cc/7BQ8-VKP7].

284 Social engineering is inherently psychologically manipulative: the online security company Kaspersky defines it as "a manipulation technique that exploits human error to gain private information, access, or valuables." *What Is Social Engineering?*, KASPERSKY, https://www.kaspersky.co.uk/resource-center/definitions/what-is-social-engineering [https://perma.cc/D8VW-SLQ3].

285 For example, Renée DiResta of the Stanford Internet Observatory has mapped the dynamics of social media "memetic propaganda" used by the "marketing agency" of the Russian IRA with canonical 1960s propaganda tactics. USENIX Enigma Conference, *supra* note 280, at 45:20.

286 Renée DiResta explains that it is now often difficult to discern which social media accounts are sockpuppet accounts operated by a U.S. adversary and which accounts are accounts operated by American individuals and organizations, because of the ongoing conversations between the two. *Id.* at 51:30.

287 *Id.* at 42:28.

288 *Cf.* David Segal, *How Bell Pottinger, P.R. Firm for Despots and Rogues, Met Its End in South Africa*, N.Y. TIMES (Feb. 4, 2018), https://www.nytimes.com/2018/02/04/business/bell-pottinger-guptas-zuma-south-africa.html [https://perma.cc/DV9D-JV8N].

tandem with the principles of PSYOP highlights the scope of a looming problem: the potential rise of a new commercial data exploitation ecosystem with fewer ethical limits that leverages known PSYOP techniques to push Internet fakery at domestic populations for profit and social disruption.

Consider the recent (unsavory) efforts[289] of a now-defunct public relations firm that was hired for the purpose of creating racial tension in South African society to distract the public's attention away from corporate clients' problematic conduct in the country.[290] Or consider the dynamics of the Facebook–Cambridge Analytica[291] relationship[292] that ultimately resulted in FTC enforcement action[293] against both companies in the United States,[294] and actions by the U.K. Information Commissioner.[295] As articulated by the FTC in its complaints against Cambridge Analytica and Facebook, Cambridge Analytica and its officers "deceived consumers by falsely claiming they did not collect any

289 As explained by one of the founders of the firm, "Morality is a job for priests . . . [n]ot P.R. men." *Id.*

290 *Id.*

291 Though Cambridge Analytica closed in 2018, key members appear to have reconstituted into a new entity also focused on data. Eli Watkins, *Cambridge Analytica Announces Closure*, CNN (May 17, 2018, 1:25 AM), https://edition.cnn.com/2018/05/02/politics/cambridge-analytica-closure/index.html [https://perma.cc/3GTW-3DLJ]; Sasha Ingber, *Cambridge Analytica Is Shutting Down After Facebook Data Controversy*, NPR (May 2, 2018, 3:24 PM), https://www.npr.org/sections/thetwo-way/2018/05/02/607782799/cambridge-analytica-is-shutting-down-after-facebook-data-controversy [https://perma.cc/C359-LHE6]; Shona Ghosh, *The Power Players Behind Cambridge Analytica Have Set Up a Mysterious New Data Company*, BUS. INSIDER (Mar. 21, 2018, 10:05 AM), https://www.businessinsider.com/cambridge-analytica-executives-and-mercer-family-launch-emerdata-2018-3 [https://perma.cc/N2Q6-CU4R].

292 Matthew Rosenberg, Nicholas Confessore & Carole Cadwalladr, *How Trump Consultants Exploited the Facebook Data of Millions*, N.Y. TIMES (Mar. 17, 2018), https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html [https://perma.cc/3QWA-2H6B].

293 The FTC ultimately levied a $5 billion fine against Facebook for its role in this deception. Press Release, FTC, FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook (July 24, 2019), https://www.ftc.gov/news-events/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions [https://perma.cc/NNS8-ELPH]. The company also experienced a one-day drop of over $100 billion dollars as the market responded to the settlement. Max A. Cherney, *Facebook Stock Drops Roughly 20%, Loses $120 Billion in Value After Warning that Revenue Growth Will Take a Hit*, MKT. WATCH (July 26, 2018, 6:59 PM), https://www.marketwatch.com/story/facebook-stock-crushed-after-revenue-user-growth-miss-2018-07-25 [https://perma.cc/J4JL-2Q2W].

294 Press Release, FTC, FTC Issues Opinion and Order Against Cambridge Analytica for Deceiving Consumers About the Collection of Facebook Data, Compliance with EU-U.S. Privacy Shield (Dec. 6, 2019), https://www.ftc.gov/news-events/press-releases/2019/12/ftc-issues-opinion-order-against-cambridge-analytica-deceiving [https://perma.cc/G44X-P8N4].

295 Natasha Lomas, *Facebook's Secret Settlement on Cambridge Analytica Gags UK Data Watchdog*, TECHCRUNCH (Jan. 26, 2021, 9:17 AM), https://techcrunch.com/2021/01/26/facebooks-secret-settlement-on-cambridge-analytica-gags-uk-data-watchdog [https://perma.cc/W4UF-CWKZ].

personally identifiable information from Facebook users who were asked to answer survey questions and share some of their Facebook profile data."[296] According to whistleblowers, Cambridge Analytica's business model[297] involved obtaining data about voters, using this to profile the voters psychologically, and then sending them targeted political ads on social media, including ads intended to discourage some types of voters from voting at all.[298] Specifically, the process employed by Cambridge Analytica appears to have merged psychographic techniques of marketers with PSYOP techniques of military professionals.[299] Indeed, the company had allegedly employed ex-government operatives, skilled professionals presumably extensively trained in PSYOP techniques and spycraft.[300] Again, former social media marketing executives voice concern: they explain that "the personality quiz that Cambridge Analytica created was nothing special"—it reflected Internet psychographic profiling already prevalent in marketing.[301] What differed was the lack of self-imposed ethical guardrails that had traditionally limited maximum exploitation of target audiences.[302]

But perhaps, even more troublingly, adversaries are increasingly comfortable with using the Internet to target civilians on domestic

---

296 The FTC also levied a $5 billion penalty against Facebook, holding it "accountable for the decisions it makes about its users' privacy as part of a settlement resolving allegations that the company violated a 2012 FTC privacy order." Press Release, FTC, FTC Sues Cambridge Analytica, Settles with Former CEO and App Developer (July 24, 2019), https://www.ftc.gov/news-events/press-releases/2019/07/ftc-sues-cambridge-analytica-settles-former-ceo-app-developer [https://perma.cc/6YY8-3769].

297 When questioning the CEO of Facebook about the Cambridge Analytica scandal, Senator Blumenthal suggested that the issue was Facebook's business model rather than Cambridge Analytica's: "Your business model is to maximize profit over privacy." Cyrus Farivar, *Facebook CEO Puts On Suit and a Smile to Try to Seduce, Assuage Senators*, ARS TECHNICA (Apr. 10, 2018, 7:11 PM), https://arstechnica.com/tech-policy/2018/04/facebook-ceo-puts-on-suit-and-a-smile-to-try-to-seduce-assuage-senators [https://perma.cc/F8W9-YFJS].

298 Janet Burns, *Whistleblower: Bannon Sought to Suppress Black Voters with Cambridge Analytica*, FORBES (May 19, 2018, 12:58 PM), https://www.forbes.com/sites/janetwburns/2018/05/19/cambridge-analytica-whistleblower-bannon-sought-to-suppress-black-voters [https://perma.cc/5JZM-6E73].

299 David A. Graham, *Not Even Cambridge Analytica Believed Its Hype*, ATLANTIC (Mar. 20, 2018), https://www.theatlantic.com/politics/archive/2018/03/cambridge-analyticas-self-own/556016 [https://perma.cc/HF4J-M6EH] ("Nix, along with colleagues Mark Turnbull and Alex Tayler, make some eye-popping claims to a reporter posing as a wealthy Sri Lankan would-be client. Turnbull speaks of engaging companies run by ex-spies from the U.K. agencies MI5 and MI6 to do research.").

300 *Id.*

301 Alexandra Samuel, *The Shady Data-Gathering Tactics Used by Cambridge Analytica Were an Open Secret to Online Marketers. I Know, Because I Was One*, VERGE (Mar. 25, 2018, 1:19 PM), https://www.theverge.com/2018/3/25/17161726/facebook-cambridge-analytica-data-online-marketers [https://perma.cc/9FDH-9R8F].

302 *Id.*

soil.[303] While the United States generally deems its own use of psychological operations to be limited to times of war, this constraint is not necessarily shared by other countries,[304] some of which have historically been more comfortable with unleashing disinformation campaigns on their own citizens.[305] Private sector digital communication tools can now be used to target not only members of the U.S. military in PSYOP operations but also their extended civilian networks of family and acquaintances. This tactic—leveraging civilian communication tools and exploiting civilians as part of PSYOP—is already in use by our adversaries in other parts of the world.[306] Indeed, the "active measures" conduct of adversaries' attempted manipulation of U.S. voters in the 2016 election might be described as a foreign PSYOP on civilians through dark PR and social media.[307]

Similarly, allegations of U.S. companies attempting to seek out government-grade surveillance technologies to deploy against their own users present concern.[308] Ostensibly, such technologies might in theory be perceived to benefit these companies' advertisers; however, in situations where some of those Internet advertisers are impersonations by foreign influence operatives, the problem of reciprocal security vulnerability arises—U.S. national security concerns cannot be divorced from commercial information security ones.[309] Because a portion of our technology fakery challenges arises from potentially

---

303 USENIX Enigma Conference, *supra* note 280, at 44:00.

304 For example, explains Renée DiResta, while Russia's IRA consisted of a privately funded group of trolls with tacit governmental approval, other efforts were operated directly through the GRU. *Id.* at 55:15.

305 Maxine David, *One Nation, One Voice: Press Control and Propaganda in Putin's Russia*, CONVERSATION (Apr. 14, 2014, 1:02 AM), https://theconversation.com/one-nation-one-voice-press-control-and-propaganda-in-putins-russia-25551 [https://perma.cc/36BP-BRRX].

306 In 2015, a Skorpion-like maneuver was carried out through private mobile phone providers, texting the families of Ukrainian soldiers false information of the soldier's death, leading family to attempt to contact the soldiers to compromise physical location through mobile phone traffic. Daniel Brown, *Russian-Backed Separatists Are Using Terrifying Text Messages to Shock Adversaries—and It's Changing the Face of Warfare*, BUS. INSIDER (Aug. 14, 2018, 5:25 PM), https://www.businessinsider.com/russians-use-creepy-text-messages-scare-ukrainians-changing-warfare-2018-8 [https://perma.cc/UK3R-TTXF]. Similarly, Ukrainian soldiers have received disinformation about alleged desertions from the Ukrainian Army—a PSYOP campaign attributed by journalists to Russia. Raphael Satter & Dmytro Vlasov, *Ukraine Soldiers Bombarded by "Pinpoint Propaganda" Texts*, AP NEWS (May 11, 2017), https://apnews.com/article/9a564a5f64e847d1a50938035ea64b8f [https://perma.cc/K4SB-HQWU].

307 S. SELECT COMM. ON INTEL., 116TH CONG., REP. ON RUSSIAN ACTIVE MEASURES CAMPAIGNS AND INTERFERENCE IN THE 2016 U.S. ELECTION VOL. 2: RUSSIA'S USE OF SOCIAL MEDIA WITH ADDITIONAL VIEWS 3 (2019), https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf [https://perma.cc/Q7GX-HZCT].

308 Coldewey, *supra* note 282.

309 As one of us has explained elsewhere, "Public sector and private sector information security concerns cannot be discretely cabined off from each other. This technical reality underpins the problem of reciprocal security vulnerability." Matwyshyn, *supra* note 268, at 1121.

foreign sources engaging in PSYOP, Internet fakery erodes national security interests, as well as commercial and civil ones.[310] Similarly, whether an Internet fakery operation is financially motivated or done at the behest of a foreign power, ultimately, the deceptive experience of the operation feels the same in real time from a user's perspective (and exploitative after the fact).[311]

Notably, the risk of creeping convergence of private sector advertising technology with military PSYOP techniques may bring to mind a warning issued by President Eisenhower in 1961. While the first part of Eisenhower's warning is familiar, the second part—the equally prescient warning against democracy becoming captive to a technological elite—is less well known. In his farewell address, President Dwight D. Eisenhower cautioned,

> [W]e must guard against the acquisition of unwarranted influence, whether sought or unsought, by the military-industrial complex. . . . We must never let the weight of this combination endanger our liberties or democratic processes. . . . Akin to, and largely responsible for the sweeping changes in our industrial-military posture, has been the technological revolution during recent decades. . . . [W]e must also be alert to the equal and opposite danger that public policy could itself become the captive of a scientific-technological elite. . . . [T]his world of ours, ever growing smaller, must avoid becoming a community of dreadful fear and hate, and be, instead, a proud confederation of mutual trust and respect.[312]

Today, we are at risk of developing the type of technological elite that may have troubled Eisenhower. Channeling Eisenhower's framing, today we face the risk of *private technology companies potentially leveraging PSYOP-trained personnel and techniques to maximally target civilian populations*, directly and indirectly giving rise to both national security and economic threats.[313] In other words, these dynamics present a risk of what might be called the rise of a *PSYOP industrial complex*.

---

310 Particularly as disinformation and misinformation have moved online, the boundaries remain unclear as to when, if ever, remote targeting of civilian populations with disinformation and misinformation crosses a line into an act of sovereignty compromise worthy of escalatory response. CATHERINE A. THEOHARY, CONG. RSCH. SERV., R45142, INFORMATION WARFARE: ISSUES FOR CONGRESS (2018).

311 Legally, these two situations would, of course, ultimately implicate different statutory regimes and trigger different particularized sanction analyses. However, at the outset of a fakery inquiry, the motivation of the faker and the connection to any broader nefarious conspiracy may not be known. Thus, a unified legal paradigm such as the one offered by this Article assists in guiding finders of fact in their analyses.

312 Eisenhower, *supra* note 253.

313 Thus, the proper level of analysis for these concerns is more than that of merely the individual user; the proper level of analysis is also that of our republic as a whole.

In summary, today's technology fakery presents a muddled scenario with blended commercial and national security implications. The next Section begins to reframe this formidable legal challenge. A blended framework becomes feasible when legal analysis starts from two unifying elements: identifying the intent behind the fakery and assessing the fakery in context.

### C. *Exploiting Vulnerabilities: Intent and Context*

*There's a sucker born every minute.*

*—likely never said by P.T. Barnum*[314]

The preceding Sections have introduced three key insights. First, prior Sections have identified four categories of Internet fakery—the dynamics of Internet MIST. Second, they have articulated two complicating dynamics—the arrival of the Internet long-con and the risk of the rise of a PSYOP industrial complex. Third, prior Sections have explained that because of the blended nature of civil and national security harms reflected in technology fakery, a robust legal paradigm must be flexible enough to conceptually encompass both. In this Section, we offer a fourth insight: two key elements vary across the most complicated instances of information fakery, regardless of whether the fakery arises from a private threat actor (e.g., a con artist) or a public one (e.g., a military information warfare unit)—the intent to deceive through fakery and the presence of affirmative acts to control the context of the fakery. Let us unpack these two elements of intent and context by analyzing the fakery dynamics of a classic children's game of Telephone[315] and a hypothetical (physical space) "Disinformation Booth."

Imagine a row of children sitting before you playing Telephone. As each child conveys the secret information to the next in a row, inevitably some scrambled results are likely to occur. Despite requests for the "Operator" and repetition,[316] a contorted result often happens due to mishearing, not due to an intentional manipulation. But imagine that a "malicious" six-year-old[317] intentionally compromises the integrity of

---

314 Glenn C. Altschuler, *P.T. Barnum Never Said "There's a Sucker Born Every Minute," but He Believed It*, PITTSBURGH POST-GAZETTE (Aug. 18, 2019, 10:00 AM), https://www.post-gazette.com/ae/books/2019/08/18/Barnum-An-American-Life-Robert-Wilson-sucker-humbug-huckster/stories/201908180008 [https://perma.cc/3A8S-LS93].

315 wikiHow Staff, *How to Play the Telephone Game*, WIKIHOW (Sept. 16, 2021), https://www.wikihow.com/Play-the-Telephone-Game [https://perma.cc/4RKW-WH5G].

316 *Id.*

317 The law generally does not consider six-year-old children capable of forming malicious intent in both criminal and contract law matters. Under states' versions of the contract law

the message. Legally cognizable harm does not result, and an "audit" of the chain of communication will readily reveal the short-statured "attacker" in the scenario. Why? Because the "threat actor" engaging in fakery is operating within the constrained context of a familiar children's game where everyone understands the context and has agreed to play.

But now compare the game of Telephone to the fakery of a con artist who serially sells fake tollbooths on the Brooklyn Bridge. The intent is deception, and victims experience the legally cognizable harm of financial loss—the cost of the "tollbooth" for the mark and the amount of "tolls" collected by the mark from third parties. The context—a bridge transit—is one that lends credibility to the con. Some bridges do, in fact, have tolls associated with them,[318] and the con artist exploits this possibility to his advantage as part of the deception. Uncontroversially, this act of fakery is one likely to lead to legal sanction. None of the "marks" understood the full context, and the perpetrator crafted content and conduct optimized to deceive.

Now consider an example in between—an orchestrated campaign of credible-looking Disinformation Booths in front of Smithsonian museums that intentionally misdirect every other person to the wrong place. The diagnosis of the problem becomes more complex, as does the articulation of a cognizable legal harm.[319] Yet, the intent to deceive remains—even if objective determination of that intent and the quantification of the harm become more difficult as a legal matter. The coordinated Smithsonian Disinformation Booth campaign might seem intuitively closer to the fake Brooklyn Bridge tollbooth than the children's game of Telephone.[320] Now imagine that the Smithsonian Disinformation Booth operation did not simply target every other patron. Imagine instead that it is operated by a militant British anti-cat/pro-bird organization, KatzRKillrz, that only targets cat owners for harm.[321] As patrons approach the Disinformation Booth, they are vetted

---

minority doctrine, children under the age of eighteen are generally entitled to set aside their contractual relationships prior to majority or shortly thereafter in most cases, other than for necessary items. For a discussion of children's intent, see, for example, Matwyshyn, *supra* note 16.

318 *See, e.g.*, Pat Ralph, *No Toll Hike on Ben Franklin, Walt Whitman or Other DRPA Bridges in 2021*, PHILLY VOICE (Dec. 10, 2020), https://www.phillyvoice.com/delaware-river-port-authority-drpa-bridges-toll-hikes-increases-2021 [https://perma.cc/G5FR-4FSA].

319 Misdirected patrons may waste time and experience physical discomfort and public humiliation due to incorrect directions in light of their particular goals. For example, consider the consequences of maliciously misdirecting a patron away from the bathroom during a potentially explosive gastrointestinal crisis.

320 The First Amendment implications of these scenarios are considered *infra* Part III.

321 Any international terrorism implications of this scenario and whether the operation may be a thinly veiled attempt to disparage the credibility of Larry as the Chief Mouser to the Cabinet Office of the United Kingdom are outside the current scope of this scenario. For a discussion of

for cat ownership using a combination of facial recognition and machine learning that performs real-time background checks for "cat-egories" of suspect purchases such as kitty litter and social media photos with cats. All patrons deemed to be cat sympathizers are then provided disinformation, directed to an area where various "cat-astrophes" await them—a "cat-tack" by an aggressive group of screaming militants, a pelting with used kitty litter, a pickpocket with sticky "paws for the cause," or a stealthy "animal owner control" vehicle that will attempt to run them over at high speed. Some targets of the Disinformation Booth may exit the experience feeling unharmed—perhaps they recognize the directions as erroneous. Some targets may rely on the information to their detriment and experience feelings of emotional distress and concern for physical safety, e.g., the scenarios with the "cat-tack." Some targets may incur small financial expense but meaningful discomfort and dignity harms, e.g., the litter-pelting scenarios. Finally, some targets may experience uncontroversially illegal criminal acts and civil harms, such as being mowed down by a vehicle. Yet, all of these harms—some more legally actionable than others—start from the same Disinformation Booth fakery.

As these hypotheticals highlight, two key points of variation are visible across instances of fakery—the degree of *intent to deceive/superior knowledge* on the part of the faker coupled with *concrete acts of context control* that lead to harms whose scope and risk are not understood by the target. Indeed, deceptive intent and high context control are hallmarks of the fakery professionals discussed in prior Sections—con artists and PSYOP professionals. But con artists and PSYOP operatives are not the only masters of fakery. The work of a third group of professional fakers also involves intentional deception and high context control—the work of professional illusionists, i.e., magicians. Magicians, like con artists and PSYOP professionals, exploit human information vulnerabilities for a living. Yet, audiences are not harmed by watching magic shows. What differentiates magicians from con artists and PSYOP professionals?

In brief, the key lies in meaningful consent and correctly placed trust. Con artists and PSYOP professionals exploit information vulnerabilities in humans without meaningful consent in context in order to further their own goals, often through abuses of trust. In contrast, magicians exploit informational vulnerabilities in furtherance of a shared enterprise of entertainment, crafting a context where risk of harm is consciously limited and without abuse of trust.

---

the tenure of Larry as Chief Mouser and its controversies, see, for example, Jill Lawless & Danica Kirka, *UK's Chief Mouser Celebrates 10 Years on the Prowl*, AP NEWS (Feb. 15, 2021), https://apnews.com/article/larry-the-cat-chief-mouser-10-years-036fa5e83422a0de41d91ab6fd91262e [https://perma.cc/496G-BGMX]. Larry's official duties include "inspecting security defences." *See Larry, Chief Mouser to the Cabinet Office*, GOV.UK, www.gov.uk/government/history/10-downing-street#larry-chief-mouser (last visited Oct. 31, 2021).

### 1. Nonconsensual Exploitation of Informational Vulnerability

*Propaganda ceases where simple dialogue begins.*

*—Jacques Ellul*[322]

As described previously, early stages in both long-con and PSYOP operations involve intentionally curating a target audience for exploitation without their meaningful consent in context.[323] This intent manifests itself in at least two concrete ways—first, in selection and targeting of the audience for maximum exploitation and, second, the nature and extent of context control measures in messaging.[324] In PSYOP, the step of identifying target audiences and mapping the audience to appropriate messaging has always been a data-intensive enterprise, incorporating both human intelligence from other information operations and insights from psychographic data.[325] In particular, the Army PSYOP Field Manual includes a discussion of "[a]dvertising and [s]ocial [m]arketing" techniques as part of psychological operations[326] and the specifics of the context of the operation.[327] Much like an Internet long-con artist or an aggressive Internet marketer might, the Field Manual advises seeking out external databases of information from private-sector sources such as "polling companies" to assist with this targeting enterprise.[328]

In crafting context control during messaging, just as a con artist might intentionally fabricate false identities or fake stories to reel in a target, psychological operations techniques also seek to create fake "stories" to control the behavior and beliefs of targets about the context. In particular, PSYOP leverages three categories of fake information—

322 JACQUES ELLUL, PROPAGANDA: THE FORMATION OF MEN'S ATTITUDES 6 (Konrad Kellen & Jean Lerner trans., Vintage Books ed. 1973).

323 "[T]he tactics, techniques, and procedures it presents should not limit creativity or imagination, provided that they adhere to Army doctrine, U.S. national policy, and the commander's intent." HEADQUARTERS, *supra* note 254, at iv.

324 *Id.* at 5-2 to 5-5.

325 *See id.* at I-10.

326 *Id.* at D-1.

327 *Id.* at 11-9. As explained by the Army PSYOP Field Manual, personnel should consider the unique context—the intended objectives of the messaging, including the selection of target audiences for greatest impact, and whether the chosen media source, themes, and techniques fit the context of the operation. *Id.*

328 *Id*. at D-2.

misinformation,[329] propaganda, and disinformation.[330] In contrast to misinformation, propaganda and disinformation display hallmarks of intentionality.[331] In other words, the distinctions among misinformation, propaganda, and disinformation turn on the objectives of the operation and, in particular, the *intent of the information disseminator*.

Returning quickly to the problems of Internet MIST from Part I, we see that intent similarly plays a key role in each of the four problematic dynamics. Manipulation and impersonation include an intent to deceive. Sequestration requires intentional acts to limit information access. Toxicity may involve the intent to harass or exploit targets.

While Bernays's theory[332] offered one possible framing for the influence of private sector branding on government communications, the work of lawyer and philosopher of technology Jacques Ellul may offer a useful framework for analysis of intent. Specifically, framed in the language of information security, Ellul's work offers potential insights about the *intent* of fakery professionals during their nonconsensual information vulnerability exploitation—a view through attackers' eyes. In his own seminal work, a book which bears the same title as Bernays's book, *Propaganda*, Ellul explains that information manipulation or "[p]ropaganda is called upon to solve problems created by technology, to play on maladjustments, and to integrate the individual into a technological world."[333]

In prior legal scholarship, Ellul's work has been applied to questions of law and medicine,[334] lobbying,[335] religion in the

---

329 According to the Army Field Manual, misinformation relates to "unintentionally incorrect information emanating from virtually anyone for reasons unknown, or to solicit a response or interest that is not political or military in origin. The recipient of this information could be anyone." *Id.* at 11-2.

330 *Id.* at 11-1 ("[D]isinformation, misinformation, propaganda, and opposing information are all being used by adversaries around the world. PSYOP personnel analyze propaganda for the purpose of determining suitable techniques for potentially countering it.").

331 Propaganda is defined by the Field Manual as "intentionally incorrect or misleading information directed against an adversary or potential adversary to disrupt or influence any sphere of national power—informational, political, military, or economic. . . . [through] attempt[ing] to mix truth and lies in a way that is imperceptible to the listener." *Id.* at 11-3.

332 *See supra* notes 294–301 and accompanying text.

333 ELLUL, *supra* note 322, at xvii.

334 *See* Charles R. DiSalvo, *Worshipping at the Altar of Technique: Manic Aggressive Medicine and Law*, 40 VILL. L. REV. 1365, 1392 (1995) (using Ellul's concept of the "transcendent over against technique" to argue that "patients, patients' families and physicians might find themselves in a culture open to those who believe that there is a more important reality than mere physical life").

335 *See* Colin Bird, *Lobbying: The Question of Propaganda*, 12 GEO. J.L. & PUB. POL'Y 451, 462–63 (2014) (using Ellul's theory of propaganda to explore the possibility "that we might consider lobbying as an element in a diffused, though no less pernicious, propaganda system of just this kind").

workplace,[336] independent judgment,[337] technology theory,[338] discourse,[339] patentability,[340] and lawyer conscience.[341] Ellul's work has been aggressively critiqued by legal scholars,[342] philosophers,[343] and media theorists for, among other reasons, its overdeterminism and superficial understanding of audience response to messaging.[344]

---

336 *See* Timothy L. Fort, *Religion in the Workplace: Mediating Religion's Good, Bad and Ugly Naturally*, 12 NOTRE DAME J.L. ETHICS & PUB. POL'Y 121, 148 (1998) (arguing that Ellul's work on the theoretical foundations of law holds an "important kernel of truth within it that business ethicists ought to consider, particularly when the demand for preciseness is most acute as it is when religious belief enters the picture").

337 Harry G. Hutchison, *Toward a Robust Conception of "Independent Judgment": Back to the Future?*, 36 U. S.F. L. REV. 335, 339 (2002) ("Ellul seems very close to the mark with his contention that human progress, not to mention human individuality, is becoming eviscerated by an increasingly technological society consumed by increasingly technical adjustments and arguments.").

338 *See* Arthur Cockfield & Jason Pridmore, *A Synthetic Theory of Law and Technology*, 8 MINN. J.L. SCI. & TECH. 475, 488 (2007) ("Dialectics, [Ellul] believed, go much further than the class struggles suggested by Marx; instead, they pervade every aspect of our lives. For Ellul, this is what arguably makes us human; our living out the tensions of life proves us to be free, to be cognitive creatures that have a full sense of agency and autonomy."); *see also* Frank Pasquale & Arthur J. Cockfield, *Beyond Instrumentalism: A Substantivist Perspective on Law, Technology, and the Digital Persona*, 2018 MICH. ST. L. REV. 821, 850 ("Ellul was concerned about what he saw as a particularly dire transition to an oppressive epoch—that of the technological society.").

339 Richard Stivers, *Technology, Discourse, and Truth*, 64 U. CIN. L. REV. 1259, 1259 (1996) ("One of Ellul's critical insights is that technology takes in more than material technology (such as machines); it includes spiritual (nonmaterial) techniques, which are either organizational or psychological, or both.").

340 Alan L. Durham, *"Useful Arts" in the Information Age*, 1999 BYU L. REV. 1419, 1451 ("[Ellul's] vision of logic and system applied to all facets of human activity suggests the *potential* scope of a broadly defined 'technology,' as well as the *potential* scope of patentable subject matter.").

341 David Barnhizer, *Princes of Darkness and Angels of Light: The Soul of the American Lawyer*, 14 NOTRE DAME J.L. ETHICS & PUB. POL'Y 371, 440 (2000) (pointing to Ellul's observation that "[t]he intelligentsia will no longer be a model, a conscience, or an animating intellectual spirit for the group. . . . They will be the servants, the most conformist imaginable, of the instruments of technique" as "what happens in regard to many lawyers in powerful law firms who. . . . become the most coopted workers in the legal profession" (alteration in original)).

342 Pasquale & Cockfield, *supra* note 338, at 852 (critiquing Ellul, stating that "[t]echnology is not given to one specific future, despite Ellul's dire warnings of a social world in which technique dominates").

343 Ellul has been critiqued, for example, for adopting an overly pessimistic position regarding the role of human agency and for spiritualist responses overriding scientific instincts. For a discussion of critiques of Ellul, see, for example, Thomas Landon Thorson, Book Review, 83 POL. SCI. Q. 117, 117 (1968) (reviewing JACQUES ELLUL, THE POLITICAL ILLUSION (Konrad Kellen trans., 1967)) (arguing Ellul's work might be read as "rather unsatisfactory" and a "hippie call to arms"); Gregory S. Butler, *The Political Moralism of Jacques Ellul*, 7 HUMANITAS 42 (1994) (explaining Ellul's "tradition of insisting upon an essentially spiritual response to the problems of modernity"). *But see* David Menninger, *Politics or Technique? A Defense of Jacques Ellul*, 14 POLITY 110 (1981).

344 Alice E. Marwick, *Why Do People Share Fake News? A Sociotechnical Model of Media Effects*, 2 GEO. L. TECH. REV. 474, 488 (2018).

However, Ellul's perspective when viewed through the lens of information security as a discussion of the *subjective* worldview held by attackers—fakery professionals such as con artists and PSYOP experts—may, nevertheless, remain useful. Thus, we offer Ellul's insights on propaganda with a limited purpose as a stepping-stone toward an objective framework: Ellul's work potentially articulates the goals of fakery creation and dissemination strategy *through the eyes of an information attacker engaged in fakery*.[345]

Specifically, Ellul explains that a "news event may be a real fact, existing objectively, or it may be only an item of information, the dissemination of a supposed fact."[346] "What makes it news is its dissemination, not its objective reality."[347] In other words, Ellul is highlighting the amplification[348] of fake information played by technology and its relationship to misinformation and disinformation—political[349] or otherwise. Ellul then argues that manipulation through information can act as a unifying force to generate potentially destructive group identity around the crafted information.[350] Speaking in words that may resonate with modern conversations about viral spread of fake news and political disinformation, Ellul explains that "[w]hat is needed, then, is continuous agitation produced artificially even when nothing in the events of the day justifies or arouses excitement,"[351]—i.e., attention control tactics.[352] This perspective rings consonant with the perspective of the Army PSYOP Field Manual and the operational choices of Russia's Information Research Agency in their informational

345 This inquiry into the information attacker's mind arguably sits apart from the empirical reality of audience perception that media experts study and the legitimate other critiques of his work. Although Ellul was writing about content creators of television and radio, his words might offer insights when considering the subjective mindset of attackers who engage in internet fakery today. *See* ELLUL, *supra* note 322. In this Article, we loosely borrow Ellul's insights within the context of a customizable media structure, one that is characterized by high levels of A/B testing and personalization, despite bearing some of the hallmarks of the aggregate dynamics Ellul raises.

346 *Id.* at 47–48.

347 *Id.* at 48.

348 *Id.*

349 *Id.*

350 Ellul's familiarity with the dynamics of Nazi social cohesion influenced his work. JACQUES ELLUL & PATRICK TROUDE-CHASTENET, JACQUES ELLUL ON RELIGION, TECHNOLOGY, AND POLITICS: CONVERSATIONS WITH PATRICK TROUDE-CHASTENET (Jacob Neusner & William S. Green eds., Joan Mendès France trans., 1998).

351 ELLUL, *supra* note 322, at 20.

352 In the context of social media and large Internet companies, Professor Tim Wu has described a portion of these dynamics in other terms through the lens of the history of commercial mass media. *See* TIM WU, THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS (2016).

interventions in the 2016 U.S. presidential election.[353] Indeed, the Internet MIST dynamics of manipulation, impersonation, sequestration, and toxicity combine to create a context akin to the one described by Ellul as ripe for exploitation by attackers: the Internet presents a technology context characterized by efforts to "create a sense of urgency"[354] and "FoMO,"[355] extreme amplification,[356] and ephemerality of memory[357] (despite existing archiving).[358]

For Ellul, master propagandists exploit faux urgency created by technology to generate what Ellul calls the "current-events man" who is a "ready target for propaganda."[359] In particular, for Ellul the context of ephemerality created by technology blends with human cognitive processing limitations and a tendency toward using emotion[360] (over thought) to create fertile territory for propagandists.[361] He explains that propagandists play on (perceived) humans' cognitive processing deficits with respect to holding conflicting ideas at the same time and identification of inconsistencies,[362] a problem he claims that can be addressed in part through dialogue.[363] He states, "To be effective, propaganda must constantly short-circuit all thought and decision. It must operate on the individual at the level of the unconscious."[364]

---

353 Emily Stewart, *Most Russian Facebook Ads Sought to Divide Americans on Race*, VOX (May 13, 2018, 12:40 PM), https://www.vox.com/policy-and-politics/2018/5/13/17349670/facebook-russia-ads-race-house-democrats [https://perma.cc/4NL7-4QWY].

354 Adam Heitzman, *How to Trigger Urgency in Your Marketing Copy*, SEARCH ENGINE J. (Apr. 21, 2021), https://www.searchenginejournal.com/create-urgency-conversions-sales/249643 [https://perma.cc/44JK-UR3U]; *How Cybercriminals Use Your Sense of Urgency Against You*, PALMETTO TECH. GRP. (Dec. 19, 2017), https://blog.goptg.com/how-cybercriminals-use-your-sense-of-urgency-against-you [https://perma.cc/BX9L-68AE].

355 Michael Hogan, *Facebook and the "Fear of Missing Out" (FoMO)*, PSYCH. TODAY (Oct. 14, 2015), https://www.psychologytoday.com/us/blog/in-one-lifespan/201510/facebook-and-the-fear-missing-out-fomo [https://perma.cc/LMS2-M2UU].

356 Molly Wood, Stephanie Hughes & Shaheen Ainpour, *How Social Media Bots Can Amplify Fake News*, MARKETPLACE (Feb. 27, 2018), https://www.marketplace.org/2018/02/27/how-social-media-bots-can-amplify-fake-news [https://perma.cc/JV5A-YH7C]; Crofton Black & Abigail Fielding-Smith, *Astroturfing, Twitterbots, Amplification—Inside the Online Influence Industry*, BUREAU INVESTIGATIVE JOURNALISM (Dec. 7, 2017), https://www.thebureauinvestigates.com/stories/2017-12-07/twitterbots [https://perma.cc/WQX7-S8Y3].

357 *Stopping Link Rot: Aiming to End a Virtual Epidemic*, NPR (Apr. 26, 2014, 10:04 AM), https://www.npr.org/sections/alltechconsidered/2014/04/26/307041846/stopping-link-rot-aiming-to-end-a-virtual-epidemic [https://perma.cc/S5BV-KH42].

358 *The Internet Never Forgets*, SCI. AM. (Aug. 18, 2008), https://www.scientificamerican.com/article/the-Internet-never-forgets [https://perma.cc/9PBU-RCZJ].

359 ELLUL, *supra* note 322, at 47.

360 *Id.*

361 *Id.*

362 *Id.*

363 *Id.* at 6.

364 *Id.* at 27.

Again, mapping onto our prior discussion of con artists' preference for more sophisticated marks as described by Smith,[365] Ellul also highlights that it is the overconfidence of more sophisticated target audiences that renders them particularly vulnerable to attackers: they assume wrongly that they cannot be manipulated.[366]

Now, let us contrast these dynamics with the work of professional illusionists—magicians. Like con artists and PSYOP operatives, magicians exploit human cognitive limitations toward a desired act of fakery.[367] However, unlike con artists and PSYOP professionals who exploit human cognitive vulnerabilities for personal gain, magicians obtain consent and craft a context that avoids audience harm and abuses of trust. Their intent is not to deceive through unfair surprise and fraud, but to entertain in a shared enterprise. These key differentiating elements—intent and context—will form the foundations for our later conversation with respect to legal interventions in Part III.

### 2. Consensual Exploitation of Informational Vulnerability

*Any sufficiently advanced technology is indistinguishable from magic.*

*—Arthur C. Clarke*[368]

In England in 1903, with flair befitting a performance artist, the famous magician and inventor Nevil Maskelyne demonstrated that the wireless telegraph was, in some ways, a fake—it was not as secure as its creator, Guglielmo Marconi, had claimed it to be.[369] In what is commonly believed in the security community to be the first documented case of the exploitation of a real-time communications technology vulnerability,[370] Maskelyne preceded Marconi's transmitted message with the words "rats, rats, rats, rats," followed by a mocking limerick[371] and cheeky Shakespearean references during the on-stage

365 "To-day your swindler goes after the most experienced and most cynical." SMITH, *supra* note 230, at 11.

366 ELLUL, *supra* note 322, at 111.

367 *See* discussion *infra* Section I.C.2.

368 ARTHUR C. CLARKE, PROFILES OF THE FUTURE 21 n.1 (1973).

369 Lauren Tousignant, *A Brief History of the Most Humiliating Hacks Ever*, N.Y. POST (Oct. 7, 2016, 12:20 PM), https://nypost.com/2016/10/07/a-brief-history-of-the-most-humiliating-hacks-ever [https://perma.cc/9LDW-QRCA].

370 From a nearby theater, he hijacked the wavelength relied upon by the Marconi wireless telegraph. Thomas McMullan, *The World's First Hack: The Telegraph and the Invention of Privacy*, GUARDIAN (July 15, 2015, 3:00 AM), https://www.theguardian.com/technology/2015/jul/15/first-hack-telegraph-invention-privacy-gchq-nsa [https://perma.cc/653S-E588].

371 *Id.* Acting with sangfroid, one of the assistants at the demonstration tore off the paper on which Maskelyne's message was printed and pocketed it, so that the audience did not notice. *See* SUNGOOK HONG, WIRELESS 110 (2001).

demonstration.[372] By exploiting a security flaw in the Marconi wireless telegraph to corrupt the integrity of Marconi's allegedly interference-proof transmission,[373] Maskelyne used one act of fakery to expose another: his exploit revealed that Marconi had attempted to trick the public into misplacing its trust in his device. Marconi had misrepresented[374] the security of his technology and its information transmission.[375]

Maskelyne's profession as a magician may appear to be a historical anachronism for the finder of a security flaw. But, in reality, it is fitting that a magician exposed an abuse of audience trust and a harmful security misrepresentation: magicians are themselves professional fakers. They are performance artists who exploit the limits of human perception and deploy "fakes"—a term of art in magic[376]—for a willing audience in a controlled setting. As the following Sections explain, illusionists offer a useful explanatory foil to con artists and PSYOP professionals in our discussion of fakery.

### a. Abracadabra—A Shared Enterprise

In a seminal article on the manipulation of attention and awareness in stage magic, renowned magicians Teller and James Randy partnered with neuroscientist coauthors to explain the techniques used by magicians to deceive.[377] In their pathbreaking interdisciplinary work, the authors argue that magicians craft cognitive and visual illusions to capitalize on shortcomings in humans' underlying neural

---

372 Paul Marks, *Dot-Dash-Diss: The Gentleman Hacker's 1903 Lulz*, NEWSCIENTIST (Dec. 20, 2011), https://www.newscientist.com/article/mg21228440-700-dot-dash-diss-the-gentleman-hackers-1903-lulz [https://perma.cc/FTR3-95E6].

373 Adrian Crenshaw, *BSidesDE 2013 1 1 110 Years of Vulnerabilities Brian Martin AKA Jericho*, YOUTUBE (Nov. 8, 2013), https://www.youtube.com/watch?v=2Z_XZn2nFWw (last visited Nov. 21, 2021). It is also perhaps the original "stunt hack." For a discussion of stunt hacking, see, for example, Mark Loveless, *Strengthening the Signal in the Noise: IoT Security and Stunt Hacking*, DECIPHER (Sept. 24, 2015), https://duo.com/decipher/strengthening-the-signal-in-the-noise-iot-security-and-stunt-hacking [https://perma.cc/2LMS-ULSF]. Some evidence also points to Maskelyne's frustrations with Marconi for his overly broad patenting conduct as part of Maskelyne's motivation for the demonstration. *See* HONG, *supra* note 371, at 104.

374 Marks, *supra* note 372 ("'I can tune my instruments so that no other instrument that is not similarly tuned can tap my messages,' Marconi boasted to London's *St. James Gazette* in February 1903.").

375 Marconi's assistant, Fleming, published an open letter seeking to identify the attacker who had engaged in "scientific hooliganism." *Id.* Maskelyne willingly identified himself, explaining that his demonstration was to protect the public from the misrepresentations of security. *Id.*

376 Laura Mallonee, *The Secret Tools Magicians Use to Fool You*, WIRED (Nov. 9, 2018, 11:00 AM), https://www.wired.com/story/magician-secret-tools-photo-gallery [https://perma.cc/5STV-NXD7]; *Fake*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/fake [https://perma.cc/WN8J-2EV6].

377 Stephen L. Macknik et al., *Attention and Awareness in Stage Magic: Turning Tricks into Research*, 9 NATURE REVS. NEUROSCIENCE 871 (2008).

mechanisms.[378] Teller explains it in this way: "Every time you perform a magic trick, you're engaging in experimental psychology. . . . I've exploited the efficiencies of your mind"[379] in much the same way that "[t]he pickpocket has found a weakness in the way we perceive motion."[380]

The authors explain that magic exists on one side of an illusion spectrum, while on the other side sit pickpockets, scammers, and con artists—professional thieves and grifters who leverage many of the same techniques as magicians.[381] In short, the authors explain that although magicians exploit and attack human cognitive weaknesses,[382] they do so for art, not crime.[383] The goal of the magician is to generate a harmless illusion with the knowledge of the spectators that it is, in fact, an illusion.[384] In contrast, the con artist or PSYOP professional generates a false sense of trust for purposes of exploitation of the target. Thus, again, herein lies the first key difference between magic when compared to con artistry and PSYOP: a shared, consensual enterprise between the faker and the targets.

The second key difference rests in the context of the deception. In magic, the "mark" is deceived (with consent) in an understood environment—the theater. The mark has full knowledge of the boundaries of the environment, understanding that the space of the theater is the playing field of the magician. Both faker and target know that the power to control the context rests with the magician. Thus, the audience expects to be deceived while in the theater, and the audience expects that the deception will end upon exiting the theater unharmed. In the context of con artistry or PSYOP, marks have no knowledge that

---

378 Or, as explained by Teller's neuroscientist coauthors, "Even when we know we're going to be tricked, we still can't see it, which suggests that magicians are fooling the mind at a very deep level." Jonah Lehrer, *Magic and the Brain: Teller Reveals the Neuroscience of Illusion*, WIRED (Apr. 20, 2009, 12:00 PM), https://www.wired.com/2009/04/ff-neuroscienceofmagic [https://perma.cc/4F93-7ADN]. "Magicians were taking advantage of these cognitive illusions long before any scientists identified them . . . ." *Id.*

379 *Id.*

380 *Id.*

381 Macknik et al., *supra* note 377, at 876.

382 Storyteller and sleight of hand master Derek DelGaudio describes these dynamics as "literally a form of self-deception." DEREK DELGAUDIO, AMORALMAN 66 (2021).

383 The audience of a magic show consents to attend the show. Members take affirmative actions signaling their consent to attend such as buying a ticket, entering the space of the magician, and voluntarily watching despite no physical limitations on their exit. Magic is a type of theatrical performance; it engages the audience with a fictional reality through particularized communication actions and tools—illusions and fakes. Macknik et al., *supra* note 377, at 871 (explaining that "[m]agic is one of the oldest and most widespread forms of performance art" and that "unlike so-called psychics, magicians do not claim to possess supernatural powers"). Indeed, other forms of performance art share these dynamics. For example, Shakespearean theater engages the audience with a fictional story through the use of iambic pentameter and period sets and costumes.

384 *Id.*

they have entered an environment controlled by a faker with superior information. Any consent on the part of the mark occurs with only the benefit of limited or manipulated information, rendering any "meeting of the minds" functionally weak and potentially legally insufficient.[385] As additional layers of technology mediate the exchange, the "mark" faces yet another diminishment of control and information imbalance—a context that may appear unfettered to the mark but, in reality, is tightly sequestered by the faker. Framed another way, audiences extend trust to a particular faker in a particular context—the key is whether the trust is misplaced.

### b. Illusions of Trust

As explained succinctly by Teller, Randi, and their coauthors, a successful illusion, much like a successful con or a successful PSYOP, is predicated on the audience's trusting their own senses and perceptions, even when those perceptions are disconnected from the reality of the situation.[386] In other words, magic combines multiple principles of human trust and perception to misdirect the audience in ways controlled by the magician.[387]

Specifically, according to Teller, Randi, and their coauthors, in illusionism, the illusion of trust is generated in the audience through triggering two forms of brain activation that neuroscience research believes to be associated with feelings of trust: activity associated with predictive ability (and feelings of conditional trust) and activity associated with social attachment (and feelings of unconditional trust).[388] Thus, we might begin to map the types of fakery to these types of trust—i.e., fakery aimed at generating a false sense of predictability and fakery aimed at generating a false sense of social attachment.[389] Magic tends to engage the first, while con artists and PSYOP professionals tend to engage the second.

In magic, the trust is conditional and limited. Fakery is done with knowledge of the marks, and members of the audience voluntarily suspend disbelief for a predetermined duration of time to jointly craft

---

385 In contract law, formation challenges are possible if one party presented the other party with knowingly false information that was the basis for a material promise in the agreement. For a discussion of formation challenges in Internet contract contexts, see, for example, Andrea M. Matwyshyn, *The Law of the Zebra*, 28 BERKELEY TECH. L.J. 155, 161 (2013).

386 *See supra* Section I.C.2.a.

387 Lehrer, *supra* note 378.

388 Brain image studies demonstrate that activation in the prefrontal cortex is essential to construction of trust and seems to be related to inferences of intention that enable prediction of other people's behavior. Macknik et al., *supra* note 377, at 876. This predictive activation connects with a maintenance of conditional trust. By comparison, unconditional trust was correlated with activity in a different piece of the brain—the septal area, which is linked to social attachment. *Id.*

389 This psycho-social distinction offers potential guidance for crafting constructive social policy interventions aimed at breaking the "spell" of a faker over his targets. Such psycho-social interventions are outside the scope of this Article.

the show with the magician. Magicians engage in accurate professional self-labeling and accurate services description.[390] Most importantly, the trust is conditional because the fakery ends when the audience leaves the theater. The fakery in magic has a hard stop. Simultaneously, any unconditional trust generated by the illusionist during the show is not abused—no unfair surprise or unexpected harms happen to the audience due to the conduct of the magician during the illusion. In contrast, a long con or a PSYOP operation might continue indefinitely harming the minds and lives of unsuspecting marks.

In summary, while magicians may exploit cognitive limitations to create trust, that trust is, ultimately, not misplaced. Thus, magicians—but not con artists or deployers of PSYOP—subscribe to the ethical principle shared by many professions of "do no harm." Building on these distinctions, *the problems of Internet fakery can be framed as problems of misplaced trust in untrustworthy people and information*. In other words, they are information security problems of *social engineering and untrustworthiness.*

The next Part explains how this concept of untrustworthiness can be connected to the computer science and philosophy understanding of "trustworthiness" to advance the fakery conversation in law.

## II. RECONSTRUCTING TRUSTWORTHINESS

> *[John Romulus Brinkley's] name stand[s] out in bold relief among the great medical luminaries of this generation.*
>
> *—John Romulus Brinkley*[391]

In Buster Keaton's 1922 film *Cops*, Buster leads a tired horse into a building to see a "goat gland specialist" named Dr. Smith; shortly thereafter, the horse exits bucking and energized.[392] That same year, a "Dr." John Romulus Brinkley in Kansas created a movie called *Rejuvenation Through Gland Transplantation*, where he chronicled his

390 For example, a magic show might start with the illusionist saying something equivalent to: "You won't believe your eyes!" The parallel types of accurate self-descriptions of identity and labeling of services do not occur with con artists and PSYOP professionals. Con artists are unlikely to introduce themselves by saying, "Hi, I'm a con artist, and I'm here to steal money from you." Similarly, people engaged in PSYOP tactics in either military or civilian contexts are unlikely to preface their created content with, "Hi, I'm a PSYOP professional, and I'm engaging in behavior intended to alter your thinking and behavior through surreptitious means."

391 JOHN ROMULUS BRINKLEY, DR. BRINKLEY'S DOCTOR BOOK intro. (1939), https://www.kansasmemory.org/item/213228/page/1 [https://perma.cc/8Z3J-NC45].

392 COPS (Joseph M. Schenck Productions 1922).

allegedly miraculously successful goat[393] testicle operations[394] that he claimed enabled thousands of men to overcome fertility issues.[395] For a mere rate of $750 per surgery,[396] Dr. Brinkley performed thousands of surgeries on patients who trusted him with their sensitive fertility issues, including a celebrity clientele that may or may not have included Huey Long, Williams Jennings Bryant, Rudolf Valentino, and President Wilson.[397] In 1923, Brinkley became a pioneer of radio by launching what was likely the most powerful radio station in the world to advertise his fertility services.[398] He used his station to "write" on-air prescriptions and then launched a network of druggists that sold his proprietary drugs.[399] In response, the American Medical Association and Federal Radio Commission, the predecessor to the FCC, targeted Brinkley for enforcement,[400] shutting down his radio station and ending his practice.[401] In reaction, Brinkley ran as a write-in candidate for governor of Kansas in 1930 and later pioneered the use of the sound truck for election messaging amplification.[402] Ultimately, Brinkley's own lawsuits resulted in his legal denouncement as a total fraud. In 1939, he sued an editor of the American Medical Association's journal for libel after the journal labeled him a medical charlatan, and, it was during that litigation that he was exposed to be, in fact, not even a licensed doctor.[403] Yet, despite this objective lack of qualifications[404] and medically questionable surgeries,[405] some of his patients swore by his

393 NUTS! (Cartuna & Gland Power Films 2016).

394 Andrew Lapin, *The Bizarre History of a Bogus Doctor Who Prescribed Goat Gonads*, NAT'L GEOGRAPHIC (July 15, 2016), https://www.nationalgeographic.com/news/2016/07/documentary-interview-medicine-science [https://perma.cc/Y8JE-CZVD].

395 Erin McCarthy, *Penny Lane on* Nuts!*, Her Documentary About "Goat Gland Doctor" John Brinkley*, MENTAL FLOSS (Jan. 22, 2016), https://www.mentalfloss.com/article/74145/penny-lane-nuts-her-documentary-about-goat-gland-doctor-john-brinkley [https://perma.cc/S6QP-HG4A].

396 *Id.*

397 *See* NUTS!, *supra* note 393, at 5:40.

398 *Id.* at 8:40.

399 *Id.* at 11:00; *see also* Penny Lane, *Footnote 076*, NOTES ON NUTS!, https://notes.nutsthefilm.com/note/076 [https://perma.cc/Y2SQ-4LU5].

400 Thayer Watkins, *John R. Brinkley and the Origin of the Radio Station Broadcasting from Del Rio, Texas*, SAN JOSÉ ST. UNIV. [https://perma.cc/Q3DQ-T59Z]. Brinkley then circumvented the law by prerecording his broadcasts. *Id.*

401 *See* NUTS!, *supra* note 393, at 16:45.

402 *Id.* at 30:47; *see also* Penny Lane, *Footnote 161*, NOTES ON NUTS!, https://notes.nutsthefilm.com/note/161 [https://perma.cc/82KM-KNZ2].

403 McCarthy, *supra* note 395.

404 William S. Powell, *Brinkley, John Romulus (Afterward Changed to Richard)*, NCPEDIA, https://www.ncpedia.org/biography/brinkley-john-romulus [https://perma.cc/ZN9V-CHW9].

405 Janet Maslin, *Fleecing the Sheep, Who Keep Coming Back for More*, N.Y. TIMES (Jan. 31, 2008), https://www.nytimes.com/2008/01/31/books/31maslin.html [https://perma.cc/NU2E-R6FE].

techniques,[406] and he enjoyed tremendous electoral support.[407] People subjectively trusted him, despite ample evidence that he was a charlatan and a con artist, objectively unworthy of their trust.

Today, instead of Brinkley's goat testicle fertility surgeries and radio prescriptions,[408] medical charlatans on the Internet promote miracle cures for autism[409] and COVID-19.[410] "Disinformation cult[s]"[411] push conspiracy theories[412] that have inspired manipulable marks to cause physical harm to themselves and to misinform others.[413] Indeed, the severity of current fakery in health information on the Internet has resulted in the World Health Organization (WHO) launching a new field of interdisciplinary study—the field of "infodemiology."[414]

The Section that follows examines questions of trust and trustworthiness, connecting them to the insights introduced in Part I—the importance of examining intent/knowledge and context. We begin by defining computer science, social science, and philosophy meanings of trust and trustworthiness.

---

406 *Id.*

407 Sam Zeff, *The Last Time a Kansas Gubernatorial Election Was This Close, It Involved Goat Glands*, KCUR (Aug. 9, 2018, 4:46 PM), https://www.kcur.org/post/last-time-kansas-gubernatorial-election-was-close-it-involved-goat-glands [https://perma.cc/YE7L-GH25].

408 *Cf. Tainted Sexual Enhancement Products*, FDA, https://www.fda.gov/drugs/medication-health-fraud/tainted-sexual-enhancement-products [https://perma.cc/J4W2-78VG].

409 Susan Scutti, *Taking "Miracle" Solution as a Cure for Autism or Cancer Is the "Same as Drinking Bleach," FDA Says*, CNN (Aug. 15, 2019, 9:44 AM), https://www.cnn.com/2019/08/15/health/bleach-miracle-cure-fda-warning/index.html [https://perma.cc/C3VH-5LF7].

410 *Fraudulent Coronavirus Disease 2019 (COVID-19) Products)*, FDA, https://www.fda.gov/consumers/health-fraud-scams/fraudulent-coronavirus-disease-2019-covid-19-products [https://perma.cc/KKS2-BHEL].

411 Violet Blue, *Anonymous Deals with its QAnon Branding Problem*, ENGADGET (Aug. 10, 2018), https://www.engadget.com/2018-08-10-anonymous-deals-with-its-qanon-branding-problem.html [https://perma.cc/E46S-KWHZ].

412 *See, e.g.*, Tim Dickinson, *How the Anti-Vaxxers Got Red-Pilled*, ROLLING STONE (Feb. 10, 2021, 8:30 AM), https://www.rollingstone.com/culture/culture-features/qanon-anti-vax-covid-vaccine-conspiracy-theory-1125197 [https://perma.cc/9HXB-GQHS].

413 EJ Dickson, *QAnon YouTubers Are Telling People to Drink Bleach to Ward Off Coronavirus*, ROLLING STONE (Jan. 29, 2020, 1:14 PM), https://www.rollingstone.com/culture/culture-news/qanon-conspiracy-theorists-coronavirus-mms-bleach-youtube-twitter-944878 [https://perma.cc/B7A2-WYZC].

414 This formalization of disinformation as an epidemic by WHO signals global concern over the contagion-like spread of Internet fakery and the severity of potential consequences of misplaced trust. *1st WHO Infodemiology Conference*, WHO, https://www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference [https://perma.cc/3YYQ-JYHK].

### A. *"Trusted" Versus "Trustworthy"*

> *The repetition of a catchword can hold analysis in fetters for fifty years and more.*
>
> —*Justice Benjamin N. Cardozo*[415]

Over a decade before Facebook's FTC consent decree violations and the Cambridge Analytica abuses,[416] the startup that would become Facebook was born in a dorm room at Harvard.[417] In a now–infamous instant message exchange in 2004, then–college student Mark Zuckerberg was asked by an acquaintance why over four thousand Harvard and other students had voluntarily shared personally identifiable information with his new social network despite no warranties about future potential (mis)uses.[418] Zuckerberg's reply was curt: "THEY 'trust me' . . . dumb fucks."[419] Putting aside Zuckerberg's pejorative analysis[420] of user intelligence (and his potentially legally questionable conduct),[421] Zuckerberg's brutal comment succinctly and accurately identified the core of all Internet fakery issues today: misplaced trust.

Legal scholars and other authors have long considered questions of trust and trustworthiness.[422] The legal scholarship reflects debate in

---

415 Charles E. Hughes, C. Sankey, W. A. Jowitt, Benjamin N. Cardozo & Frederick Pollock, *Mr. Justice Holmes*, 44 HARV. L. REV. 677, 689 (1931).

416 *See supra* text accompanying notes 296–301.

417 Facebook was founded in 2004. Nicholas Carlson, *At Last—The Full Story of How Facebook Was Founded*, BUS. INSIDER (Mar. 5, 2010, 4:10 AM), https://www.businessinsider.com/how-facebook-was-founded-2010-3 [https://perma.cc/C8CC-VQA4].

418 *If Facebook Will Not Fix Itself, Will Congress?*, ECONOMIST (Apr. 14, 2018), https://www.economist.com/united-states/2018/04/11/if-facebook-will-not-fix-itself-will-congress [https://perma.cc/L6AW-QAKZ].

419 *Id.*

420 The inspiration for Facebook, originally known as The Facebook, arose from Zuckerberg's observation that picture book photos of some of his college classmates were "horrendous" and were not necessarily more attractive than "farm animals." Claire Hoffman, *The Battle for Facebook*, ROLLING STONE (Sept. 15, 2010, 4:45 PM), https://www.rollingstone.com/culture/culture-news/the-battle-for-facebook-242989 [https://perma.cc/5WTE-QV92].

421 Harvard raised objections to his alleged acts of copyright infringement and transgressive security conduct, and Zuckerberg faced potential expulsion for his conduct. Katharine A. Kaplan, *Facemash Creator Survives Ad Board*, HARV. CRIMSON (Nov. 19, 2003), https://www.thecrimson.com/article/2003/11/19/facemash-creator-survives-ad-board-the [https://perma.cc/T5F8-HUDY].

422 *See, e.g.*, Frank B. Cross, *Law and Trust*, 93 GEO. L.J. 1457 (2005).

the context of reputation,[423] network neutrality,[424] electronic commerce,[425] online interactions generally,[426] competition,[427] social commerce,[428] contextual factors,[429] risk management,[430] privacy,[431] trademark law,[432] consumer autonomy,[433] criminal law,[434] form contracting,[435] campaign finance,[436] the Internet of Things,[437] self-

---

423 Adam Thierer, Christopher Koopman, Anne Hobson & Chris Kuiper, *How the Internet, the Sharing Economy, and Reputational Feedback Mechanisms Solve the "Lemons Problem*," 70 U. MIA. L. REV. 830 (2016).

424 A trustworthy system has been described as one that "does what people expect it to do—and not something else—despite environmental disruption, human user and operator errors, and attacks by hostile parties." Trustworthiness is a "multidimensional" concept encompassing "correctness, reliability, security . . . privacy, safety, and survivability." Security, in turn, means resistance to attacks that "can compromise the secrecy, integrity, or availability of data and services."

Aaron J. Burstein & Fred B. Schneider, *Trustworthiness as a Limitation on Network Neutrality*, 61 FED. COMMC'NS L.J. 591, 594 (2009) (quoting COMM. ON INFO. SYS. TRUSTWORTHINESS, COMPUT. SCI. & TELECOMMS. BD., COMM'N ON PHYSICAL SCIS., MATHEMATICS & APPLICATIONS & NAT'L RSCH. COUNCIL, TRUST IN CYBERSPACE 13–14 (Fred B. Schneider ed., 1999)) (discussing trustworthiness in the context of network neutrality).

425 *See, e.g.*, Susan Block-Lieb, *E-Reputation: Building Trust in Electronic Commerce*, 62 LA. L. REV. 1199 (2002).

426 *See, e.g.*, Justin (Gus) Hurwitz, *Trust and Online Interaction*, 161 U. PA. L. REV. 1579 (2013).

427 *See, e.g.*, Dan Jerker B. Svantesson, *The Extraterritoriality of EU Data Privacy Law—Its Theoretical Justification and Its Practical Effect on U.S. Businesses*, 50 STAN. J. INT'L L. 53, 57 (2014).

428 *See, e.g.*, Julia Y. Lee, *Trust and Social Commerce*, 77 U. PITT. L. REV. 137 (2015).

429 *See, e.g.*, Helen Nissenbaum, *Securing Trust Online: Wisdom or Oxymoron?*, 81 B.U. L. REV. 635 (2001).

430 *See, e.g.*, Paul Slovic, *Trust, Emotion, Sex, Politics, and Science: Surveying the Risk Assessment Battlefield*, 1997 U. CHI. LEGAL F. 59.

431 *See, e.g.*, Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431 (2016); Ari Ezra Waldman, *Privacy as Trust: Sharing Personal Information in a Networked World*, 69 U. MIA. L. REV. 559 (2015).

432 *See, e.g.*, Ariel Katz, *Beyond Search Costs: The Linguistic and Trust Functions of Trademarks*, 2010 BYU L. REV. 1555.

433 *See, e.g.*, Jay P. Kesan, Carol M. Hayes & Masooda N. Bashir, *A Comprehensive Empirical Study of Data Privacy, Trust, and Consumer Autonomy*, 91 IND. L.J. 267, 347 (2016).

434 *See, e.g.*, Rosann Greenspan, *Gaining Public Trust in the Criminal Legal Process*, 66 S. CAL. L. REV. 2199 (1993); Lauren M. Ouziel, *Legitimacy and Federal Criminal Enforcement Power*, 123 YALE L.J. 2236 (2014).

435 *See, e.g.*, P. Göran T. Hägg, *The Economics of Trust, Trust-Sensitive Contracts, and Regulation*, 14 INT'L REV. L. & ECON. 437 (1994); Eli Bukspan, *The Notion of Trust as a Comprehensive Theory of Contract and Corporate Law: A New Approach to the Conception that the Corporation Is a Nexus of Contract*, 2 HASTINGS BUS. L.J. 229 (2006).

436 *See, e.g.*, Lawrence Lessig, *A Reply to Professor Hasen*, 126 HARV. L. REV. F. 61 (2013).

437 *See, e.g.*, Julie Brill, *The Internet of Things: Building Trust and Maximizing Benefits Through Consumer Control*, 83 FORDHAM L. REV. 205 (2014).

help,[438] reputation,[439] personhood,[440] cryptocurrency,[441] corporate boards[442] and corporate law more generally,[443] expert witness testimony,[444] exceptions to the hearsay rule,[445] confidence in the judiciary,[446] medical treatment,[447] the Confrontation Clause,[448] human subjects research,[449] community norms,[450] and the Sixth Amendment generally.[451] Yet, in all but one of these law review articles,[452] the framings of "trustworthiness" adopted by legal scholars conflict with the understanding of the concept from computer science. No prior legal scholarship rigorously bridges computing and non-computing discussion of trustworthiness in the context of technology fakery. Ergo, let us embark on precisely such an analysis.

---

438 *See, e.g.*, D. James Greiner, Dalié Jiménez & Lois R. Lupica, *Self-Help, Reimagined*, 92 IND. L.J. 1119 (2017).

439 *See, e.g.*, Avner Ben-Ner & Louis Putterman, *Trusting and Trustworthiness*, 81 B.U. L. REV. 523 (2001).

440 *See, e.g.*, Sonia M. Suter, *Disentangling Privacy from Property: Toward a Deeper Understanding of Genetic Privacy*, 72 GEO. WASH. L. REV. 737 (2004); Adam B. Seligman, *Role Complexity, Risk, and the Emergence of Trust*, 81 B.U. L. REV. 619 (2001).

441 *See, e.g.*, Rebecca M. Bratspies, *Cryptocurrency and the Myth of the Trustless Transaction*, 25 MICH. TECH. L. REV. 1 (2018).

442 *See, e.g.*, Joan MacLeod Heminway, *Sex, Trust, and Corporate Boards*, 18 HASTINGS WOMEN'S L.J. 173 (2007).

443 *See, e.g.*, Margaret M. Blair & Lynn A. Stout, *Trust, Trustworthiness, and the Behavioral Foundations of Corporate Law*, 149 U. PA. L. REV. 1735 (2001).

444 *See, e.g.*, Julie A. Seaman, *Triangulating Testimonial Hearsay: The Constitutional Boundaries of Expert Opinion Testimony*, 96 GEO. L.J. 827 (2008).

445 *See, e.g.*, James Joseph Duane, *The Four Greatest Myths About Summary Judgment*, 52 WASH. & LEE L. REV. 1523 (1995).

446 *See, e.g.*, Zelda M. DeBoyes, *Public Trust: Past, Present, Future*, 52 No. 2 JUDGES' J. 8, 9 (2013).

447 *See, e.g.*, Mark A. Hall, *Caring, Curing, and Trust: A Response to Gatter*, 39 WAKE FOREST L. REV. 447 (2004); M. Gregg Bloche, *Trust and Betrayal in the Medical Marketplace*, 55 STAN. L. REV. 919 (2002).

448 *See, e.g.*, Seaman, *supra* note 444.

449 *See, e.g.*, Chao-Tien Chang, *Bank on We the People: Why and How Public Engagement Is Relevant to Biobanking*, 25 MICH. TECH. L. REV. 239 (2019).

450 *See, e.g.*, Lan Cao, *Looking at Communities and Markets*, 74 NOTRE DAME L. REV. 841 (1999).

451 *See, e.g.*, *Sentencing Guidelines*, 38 GEO. L.J. ANN. REV. CRIM. PROC. 681 (2009).

452 *See* Burstein & Schneider, *supra* note 424.

### 1. Trustworthiness and Computing

In January 2002, Bill Gates sent an email to all Microsoft employees[453] launching Microsoft's Trustworthy Computing Initiative.[454] In hindsight, this initiative would become a critical transformational milestone in both the life of Microsoft Corporation and the broader technology industry.[455] Gates's email announced that "Trustworthy Computing is the highest priority for all the work we are doing. We must lead the industry to a whole new level of Trustworthiness in computing. . . . Trustworthy Computing is computing that is as available, reliable and secure as electricity, water services and telephony."[456] In other words, Gates's email highlighted a key computing term of art as the lodestar for the company's future: trustworthiness.

Computer science draws a distinction between the concepts of "trusted" and "trustworthy." In computer science, the word "trusted" refers to technical trust—meaning which components of the system rely on other components of the system as a technical matter—reasonably or unreasonably. [457] In other words, saying a component or system is trusted merely signals a dependency[458] as a descriptive matter; it is not an attestation of error-free functionality. It is merely a technical description, not an assertion of code or system quality. As explained by Professor Ross Anderson, "[T]he 'trusted computing base' is the set of all hardware, software and procedural components that enforce the security policy" and "a trusted component is one which can break security."[459] No normative question of appropriateness of trust is included in this framing; it merely describes the functional relationship

---

453 Bill Gates, *Bill Gates: Trustworthy Computing*, WIRED (Jan. 17, 2002, 12:00 PM), https://www.wired.com/2002/01/bill-gates-trustworthy-computing [https://perma.cc/6DA8-NH6M].

454 CRAIG MUNDIE, PIERRE DE VRIES, PETER HAYNES & MATT CORWINE, MICROSOFT CORP., TRUSTWORTHY COMPUTING 3–4 (2002), http://download.microsoft.com/documents/australia/about/trustworthy_comp.doc [https://perma.cc/G7V4-BTHV].

455 While this laudable leadership on setting a culture of security may have arisen sua sponte, it also coincided with an ongoing FTC investigation into the security and privacy practices of Microsoft Passport, after a July 2001 complaint from a coalition of consumer groups. Press Release, FTC, Microsoft Settles FTC Charges Alleging False Security and Privacy Promises (Aug. 8, 2002), https://www.ftc.gov/news-events/press-releases/2002/08/microsoft-settles-ftc-charges-alleging-false-security-privacy [https://perma.cc/LC5F-A4KD].

456 Gates, *supra* note 453.

457 Ross Anderson, *The Trusted Computing Base* (Jan. 12, 1996, 10:49 AM) [https://perma.cc/NYQ9-KEFJ].

458 For a discussion of dependencies in computing, see, for example, Rebecca Elizabeth Grinter, Understanding Dependencies: A Study of the Coordination Challenges in Software Development (1996) (Ph.D. dissertation, University of California, Irvine), https://www.cc.gatech.edu/~beki/t1.pdf [https://perma.cc/M8S9-E2QA].

459 Anderson, *supra* note 457.

of components with respect to potential compromise.[460] In other words, in computing, a component or system is "trusted" when, in fact, other parts of the system rely on it, even potentially to their detriment. Indeed, the adjective is sometimes used for processes that are not subjected to checks for security or correct operation, or applications that are given higher levels of system access,[461] because it is simply assumed[462] (potentially incorrectly) that they will behave as they should.[463] Thus, a trusted component or system may also be an abusive one unworthy of trust, in the sense understood in fields outside of computer science.[464] In other words, trusted systems might not be trust*worthy* systems.

In contrast to the term "trusted," the use of the word "trustworthy" signals a promise that a user's trust is correctly placed—that the components or system reflect objectively testable properties of proper functionality in context.[465] The National Institute of Standards and Technology (NIST) has historically defined trustworthy systems as ones that are "capable of operating within defined levels of risk despite the environmental disruptions, human errors, structural failures, and purposeful attacks that are expected to occur in its environment of

---

460 "In the commonplace use of language, when we say that we trust someone we mean that we rely on that person to do—or not to do—certain things." *Id.*

461 In the .Net framework, for example, applications with higher trust levels are permitted to access more resources. *Security Trust Levels in Accessing Resources*, MICROSOFT (Sept. 15, 2021), https://docs.microsoft.com/en-us/dotnet/framework/data/transactions/security-trust-levels-in-accessing-resources [https://perma.cc/3KZ7-PXUE].

462 Sometimes this belief is unjustified. The *trusted computing base* of a computing system consists of all the elements of the system—hardware, firmware, and software—that together enforce a security policy, as implemented by personnel. *See* U.S. DEP'T OF DEF., DOD 5200.28-STD, DEPARTMENT OF DEFENSE STANDARD: DEPARTMENT OF DEFENSE TRUSTED COMPUTER SYSTEM EVALUATION CRITERIA 112 (1985), https://csrc.nist.gov/csrc/media/publications/conference-paper/1998/10/08/proceedings-of-the-21st-nissc-1998/documents/early-cs-papers/dod85.pdf [https://perma.cc/6N2T-9KH2].

463 For instance, in the language D, the function mark "@trusted" allows compiler checks for memory safety to be bypassed. Steven Schveighoffer, *How to Write @trusted Code in D*, D BLOG (Sept. 28, 2016), https://dlang.org/blog/2016/09/28/how-to-write-trusted-code-in-d [https://perma.cc/5CVE-DKTK].

464 *See infra* text accompanying notes 493–502.

465 Specifically, the National Institute of Standards and Technology defines trustworthy systems as "[c]omputer hardware, software and procedures that: (1) are reasonably secure from intrusion and misuse; (2) provide a reasonable level of availability, reliability, and correct operation; (3) are reasonably suited to performing their intended functions; and (4) adhere to generally accepted security procedures." D. RICHARD KUHN, VINCENT C. HU, W. TIMOTHY POLK & SHU-JEN CHANG, NAT'L INST. OF STANDARDS & TECH., NIST SPECIAL PUB. 800-32, INTRODUCTION TO PUBLIC KEY TECHNOLOGY AND THE FEDERAL PKI INFRASTRUCTURE 52 (2001), https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-32.pdf [https://perma.cc/W4MN-WH9R]; *see also Trustworthiness*, NAT'L INST. OF STANDARDS AND TECH., https://csrc.nist.gov/glossary/term/trustworthiness [https://perma.cc/GY4W-CK7N].

operation."[466] In other words, trustworthiness means that a component or system can be expected to successfully perform in particular, stipulated contexts. It is a testable assertion. Similarly, the description of a system as "trusted" in the technical sense, or as "trustworthy," usually describes the system's properties through the eyes of the system's builders based on their own in-house testing.[467] Thus, assertions of trustworthiness should be understood as a representation and warranty of quality in functionality in the contract law sense. Nevertheless, the description of the builder does not necessarily accurately describe the system as an objective matter, such as through the eyes of users or of a neutral auditor; the need for independent technical validation of trustworthiness assertions remains.

In summary, deployed trustworthy systems are trusted[468] but *not all trusted systems are trustworthy* in a technical sense. Thus, whether a technologist describes a system as merely "trusted" or as "trustworthy" is an important distinction, and it is one that may be missed by policymakers.[469] Yet, it conceptually sits at the heart of the inquiries they are conducting.[470]

Now let us merge these technical discussions of trust and trustworthiness with social and philosophical ones about the human experience of trustworthiness.

---

466 JOINT TASK FORCE TRANSFORMATION INITIATIVE, NAT'L INST. OF STANDARDS & TECH., NIST SPECIAL PUB. 800-53, SECURITY AND PRIVACY CONTROLS FOR FEDERAL INFORMATION SYSTEMS AND ORGANIZATIONS B-25 (2013), https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf [https://perma.cc/99P9-GQ36].

467 In other words, while the builder of a system might assert its trustworthiness, until such claims are vetted by a neutral third party, its accuracy is unknown.

468 The exception would be a deployed system so straightforward as to have no subsidiary components trusted by other components—a highly unusual situation.

469 In policy conversations on topics of Internet fakery, exchanges between members of Congress and tech company executives have included discussions regarding "trust." Theodore Schleifer, *Congress Couldn't Agree on What Exactly Was Wrong with Mark Zuckerberg. But They All Wanted a Piece of Him*, VOX (Oct. 23, 2019, 4:40 PM), https://www.vox.com/recode/2019/10/23/20928859/libra-hearing-congress-mark-zuckerberg [https://perma.cc/S8W2-A6FY].

470 Because of the critical distinction of trustworthiness versus trust, exchanges between technology executives and policymakers can devolve into miscommunications, at best. At worst, these exchanges create an opportunity for technologists to potentially game congressional inquiries through wordsmithed technical responses. A statement such as, "We certainly have work to do to build trust" can be understood to mean increasing internal technical dependency or building users' belief that a system is trustworthy—a public relations framing of perception management, rather than a promise of a rigorous technical self-examination of technical information security and trustworthiness. In other words, committing to building user trust is not the same thing as committing to building a system that is more trustworthy. *Mark Zuckerberg Tried to Charm Congress. He Got Slammed.*, N.Y. TIMES (Oct. 25, 2019), https://www.nytimes.com/2019/10/24/business/dealbook/mark-zuckerberg-facebook-libra.html [https://perma.cc/KE7C-S2RW]. It therefore does not necessarily mean ensuring confidentiality, integrity, and availability for the users who rely on the functionality of the system. The words that would signal this concept of building a technical system where reliance is justified would be, "We have work to do on building trustworthiness."

### 2. Trustworthiness and Society

When asking the question of how human beings trust,[471] philosophers of trust have synthesized the subject into three targets or types of trust—trust in objects, trust in people, and trust in institutions.[472] In a meta-analysis of the academic literatures of trust,[473] Professor Katherine Hawley explains that three ideas consistently span disciplines in analyses of trust.[474] First, trust in people and institutions requires a richer analysis than the identification of trust in objects.[475] Second, trust in people and trust in institutions involve expectations of good intentions and competence in context; either one standing alone is not enough to support trust.[476] In other words, the trust literature reinforces the two key elements identified by our analysis of con artistry—intent and context. Third, the philosophy of trust adds a critical new consideration: the existence of *distrust* as an independent construct.[477] Thus, explains Hawley, distrust does not equal an absence of trust; instead it exists as an independent, equally potent opposing force.[478] The absence of trust is more accurately considered a state of uncertainty that could fluctuate either in favor of trust or in favor of distrust.[479]

In brief, Professor Hawley explains that trust equals an expectation of honesty and knowledge in context, while distrust equals the situation where honesty or knowledge is doubted.[480] She explains that the analysis, consequently, turns on an assessment of the commitment of the party asking for trust, as well as the extent of their knowledge in context; in other words, an assessment of their trustworthiness.[481] Professor Hawley connects trust to the state of trustworthiness by explaining that preemptively granting trust does not always result in trustworthiness and that humans can exploit the vulnerability of preemptive trust extension.[482]

Indeed, Professor Hawley reminds us that while some people can be trustworthy in one context, they may be inherently untrustworthy in

---

471 Here trust is used in the colloquial sense, not in the technical computer science sense—meaning that trust in this context comes with an implicit expectation of trustworthiness.

472 *See, e.g.*, KATHERINE HAWLEY, TRUST: A VERY SHORT INTRODUCTION (2012).

473 *Id.* at 11–12.

474 *Id.* at 5.

475 *Id.* at 98.

476 *Id.* at 6–7, 46–47.

477 *Id.* at 8.

478 *Id.*

479 *Id.*

480 *Id.*

481 *Id.* at 51.

482 *Id.* at 35.

another.[483] Assessments of trustworthiness depend in part on the extent of possible damage and its correctability. Trust is a benefit only if properly directed; if inappropriately directed it generates the impression that the "truster" is gullible, naïve, or irresponsible. Thus, a personal sanction of sorts occurs when trust is misplaced—a "penalty" for a failed determination of trustworthiness.[484] Here, let us briefly reconnect this discussion with the discussion from prior Sections. Using words that might be applied to the dynamics of art forgeries[485] and also to the dynamics of modern Internet fakery,[486] confidence artist Edward Smith explained that successful manipulation of a "mark" turns on tricking the mark into two types of incorrect trustworthiness determinations. The first involves exploiting the trust of the mark directly; the second involves the mark's functional self-exploitation as they struggle to project their own trustworthiness to third parties. Smith explained that the need of marks to preserve their own trusted roles within their social and professional networks[487] makes them in part vulnerable to con artistry.[488]

In summary, the philosophy and social science literatures of trust connect to the computer science literatures of trust through the concept of trustworthiness. Literature from both computing and other fields highlights that successful relationships—whether technical or human—involve correct determinations of trustworthiness, i.e., a determination of good intentions and relevant knowledge to succeed in a particular context. Using the insight that trust and distrust are opposite ends of a sliding scale, we argue that trustworthiness and untrustworthiness similarly sit on opposite ends of a scale. Thus, *untrustworthiness*—the opposite condition to trustworthiness—refers to the condition where trust is misplaced in code and/or other people.

Conceptually, numerous areas of law rely on the idea that trustworthiness can be assumed by default in the absence of a warning to the contrary. In tort, product liability presumes that goods are fit for purpose and trustworthy for use as the manufacturer has described.[489] Contract law presumes that a party who formally memorializes a set of

---

483 *Id.* at 65.

484 *Id.* at 14–15. Conversely, being miserly with preemptive trust means potentially missing out on positive opportunities. Thus, Internet fakery is problematic not only because it can result in people trusting things they should not, but also because it can result in people becoming afraid to trust things that are, in fact, trustworthy.

485 *See supra* Section I.A.1.b.

486 *See* discussion of influencer manipulation *supra* Section I.A.1.b.

487 SMITH, *supra* note 230, at 17, 19.

488 *See, e.g.*, Jeff Maysh, *The Man Who Sold the Eiffel Tower. Twice.*, SMITHSONIAN MAG. (Mar. 9, 2016), https://www.smithsonianmag.com/history/man-who-sold-eiffel-tower-twice-180958370 [https://perma.cc/J2S4-UBE3] (discussing marks' failures to report the swindle to police to avoid public ridicule).

489 *See, e.g.*, Stuart M. Speiser, Charles F. Krause, Alfred W. Gans & Monique C. M. Leahy, *Fitness for Ordinary Purposes*, 6 AM. L. TORTS § 18:53 (2021).

representations and warranties should be presumed trustworthy until demonstrated otherwise.[490] Regimes such as securities regulation[491] and medical device regulation rely on voluntary disclosures which by default are presumed trustworthy until proven otherwise.[492] As such, the law's focus on sanction implicitly already focuses on situations where untrustworthiness has given rise to harm—a physical harm, a material breach, a material nondisclosure, and the like. So too *in the context of technology fakery, the law should primarily focus on articulating recourse for situations arising from problems of untrustworthiness.*

But before we embark upon crafting a framework focused on untrustworthiness and fakery, let us first briefly understand the boundaries of any such legal approaches—the constraints and protections of the First Amendment in the context of fakery.

### B. *Freedom from Untrustworthy Content and Conduct*

*Father, I cannot tell a lie.*

*—maybe said by George Washington*[493]

Prior Sections of this Article have articulated the elements of intent and context as the dispositive variables in determining untrustworthiness. This Section asks, "To what extent can untrustworthiness of Internet fakery be regulated?" What baseline constitutional constraints exist on approaches that aim to minimize content and conduct untrustworthiness?

The Commerce Clause specifically empowers Congress to regulate commerce between the states,[494] and the Supreme Court has expansively understood this "regulability" to include the Internet.[495] Ergo, perhaps the most obvious constraint on congressional and state regulatory power over Internet fakery exists through the First Amendment. To date, two main categories of regulatory approaches have been

---

490 *See, e.g.*, John E. Flanagan, *The Duty of Good Faith in Contracts: Mutual Expectations Set the Parameters*, 70 WIS. LAW. (1997).

491 *See, e.g.*, *Compliance with the Periodic Reporting Requirements of the Securities Exchange Act of 1934*, MOD. CORP. CHECKLISTS § 15:8 (2021).

492 *See, e.g.*, Charles J. Nagy, Jr., *Devices Marketed with Premarket Approval*, AM. L. PRODS. LIAB. 3D § 91:20 (2021).

493 *George Washington and the Cherry Tree*, NAT'L PARK SERV., https://www.nps.gov/articles/george-washington-and-the-cherry-tree.htm [https://perma.cc/W6LF-AS66].

494 U.S. CONST. art. I, § 8, cl. 3; Champion v. Ames, 188 U.S. 321 (1903); Brooks v. United States, 267 U.S. 432 (1925); New York v. United States, 505 U.S. 144 (1992); Printz v. United States, 521 U.S. 898 (1997).

495 *See, e.g.*, Reno v. ACLU, 521 U.S. 844 (1997).

proposed—false content prohibitions[496] and various types of content-neutral approaches. These proposed or possible content-neutral approaches have included various forms of mandatory disclosure and labeling requirements,[497] restrictions on amplification of certain content,[498] altered moderation liability in various forms,[499] information reuse restrictions,[500] and personalization restrictions.[501] First Amendment concerns related to each of the approaches will be introduced in this Section but discussed in greater detail in the companion article to this piece, *Superspreaders*.[502]

### 1. False Content Prohibitions in Context

Perhaps the most obvious First Amendment question involves the extent to which false speech is regulable in Internet contexts. A common misunderstanding exists in the popular press[503] and perhaps even in some corners of the legal academy that *United States v. Alvarez* in essence created a First Amendment "right to lie."[504] It did not.

In *Alvarez*, the Court explained that a generalized prohibition on all false speech would present a new category of restriction[505] that is content-based and without adequate historical basis[506] and it cautioned

---

496 Deceptive Practices and Voter Intimidation Prevention Act of 2019, S. 1834, 116th Cong. § 3 (2019).

497 Online Privacy Act of 2019, H.R. 4978, 116th Cong. § 107 (2019).

498 Protecting Americans from Dangerous Algorithms Act, H.R. 8636, 116th Cong. (2020).

499 SAFE TECH Act, S. 299, 117th Cong. (2021); *see also* Communications Decency Act, 47 U.S.C. § 230.

500 H.R. 4978.

501 Algorithmic Justice and Online Platform Transparency Act, S. 1896, 117th Cong. (2021).

502 Draft on file with authors.

503 Editorial Board, *The Supreme Court Defends the Right to Lie*, WASH. POST (June 29, 2012), https://www.washingtonpost.com/opinions/the-supreme-court-defends-the-right-to-lie/2012/06/29/gJQAmC2RCW_story.html [https://perma.cc/3WBT-V2D2].

504 567 U.S. 709 (2012). *But see, e.g.*, Blitz, *supra* note 127.

505 *Alvarez*, 567 U.S. at 722 ("[T]he Court has acknowledged that perhaps there exist 'some categories of speech that have been historically unprotected . . . but have not yet been specifically identified or discussed . . . in our case law.' . . . The Government has not demonstrated that false statements generally should constitute a new category of unprotected speech on this basis." (internal citations omitted)).

506 *Id.* at 717–18 ("[C]ontent-based restrictions on speech have been permitted, as a general matter, only when confined to the few 'historic and traditional categories [of expression] long familiar to the bar.' Among these categories are advocacy intended, and likely, to incite imminent lawless action, obscenity, defamation, speech integral to criminal conduct, so-called 'fighting words,' child pornography, fraud, true threats, and speech presenting some grave and imminent threat the government has the power to prevent. . . . These categories have a historical foundation in the Court's free speech tradition. . . . Absent from those few categories where the law allows content-based regulation of speech is any general exception to the First Amendment for false statements." (second alteration in original) (internal citations omitted)).

against "free-wheeling" approaches[507] and tests involving relative balancing of social costs and benefits.[508] But, the Court also explained that "there are instances in which the falsity of speech bears upon whether it is protected. Some false speech may be prohibited even if analogous true speech could not be. This opinion does not imply that any of these targeted prohibitions are somehow vulnerable."[509] The Court clarified that although some false speech in public debate[510] is inevitable,[511] false statements can present systemic risk in key situations where core governmental or market functions are impaired. Ergo, it is not falsity alone that gives rise to the restriction[512] in, for example, situations of fraud, impersonation,[513] perjury,[514] false statements to government officials,[515] or defamation.[516] Citing to *Virginia State Board of Pharmacy v. Virginia Citizens Consumer Council, Inc.*, the Court continued, "Where false claims are made to effect a fraud or secure moneys or other valuable considerations, say offers of employment, it is well established that the Government may restrict speech without affronting the First Amendment."[517]

---

507 *Id.* at 724 ("In assessing content-based restrictions on protected speech, the Court has not adopted a free-wheeling approach . . . ." (internal citations omitted)).

508 *Id.* at 717 ("[T]his Court has rejected as 'startling and dangerous' a 'free-floating test for First Amendment coverage . . . [based on] an ad hoc balancing of relative social costs and benefits.'" (alterations in original) (citation omitted)).

509 *Id.* at 721.

510 *Id.* at 723 (referencing George Orwell's *1984* and stating that "[o]ur constitutional tradition stands against the idea that we need Oceania's Ministry of Truth").

511 *Id.* at 718 ("[T]he common understanding that some false statements are inevitable if there is to be an open and vigorous expression of views in public and private conversation, expression the First Amendment seeks to guarantee." (internal citations omitted)).

512 *Id.* at 719 ("These quotations all derive from cases discussing defamation, fraud, or some other legally cognizable harm associated with a false statement . . . .").

513 *Id.* at 721 ("Statutes that prohibit falsely representing that one is speaking on behalf of the Government, or that prohibit impersonating a Government officer, also protect the integrity of Government processes, quite apart from merely restricting false speech.").

514 *Id.* at 720–21 ("It is not simply because perjured statements are false that they lack First Amendment protection. Perjured testimony 'is at war with justice' because it can cause a court to render 'a judgment not resting on truth.' Perjury undermines the function and province of the law and threatens the integrity of judgments that are the basis of the legal system." (quoting *In re* Michael, 326 U.S. 224, 227 (1945))).

515 *Id.* at 720 ("Section 1001's prohibition on false statements made to Government officials, in communications concerning official matters, does not lead to the broader proposition that false statements are unprotected when made to any person, at any time, in any context.").

516 *Id.* at 719 ("[W]hen considering some instances of defamation and fraud, moreover, the Court has been careful to instruct that falsity alone may not suffice to bring the speech outside the First Amendment. The statement must be a knowing or reckless falsehood.").

517 *Id.* at 723 (citing Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, Inc., 425 U.S. 748, 771 (1976)).

Turning to the particular facts of the case, where an individual was prosecuted under the Stolen Valor Act for lying about a military honor at a public meeting, the Court can be read to perform a three-part analysis. First, the Court examined the legal nature of the false information.[518] In analyzing the legal nature of the defendant's fakery, the Court noted that the Stolen Valor Act lacked nuance: it failed to differentiate between public statements of untrustworthy information and those made in private or for creative purposes.[519]

Second, the Court performed an implicit assessment of the intent and knowledge of the disseminator and an explicit assessment of the nature of the alleged harm.[520] The Court explained that for the defendant, "[l]ying was his habit,"[521] and the Court viewed the defendant's speech as merely an attempt at social self-aggrandizement.[522] The necessity[523] of a falsity prohibition[524] and a causal link to a recognized category of harm were not sufficiently demonstrated.[525] Indeed, Professor Rodney Smolla explains that in *Alvarez*, the particular drafting was dispositive, resulting in a situation where "the law could not survive 'exacting' scrutiny."[526] He notes that the "plurality heavily emphasized that [t]he First Amendment requires that the Government's chosen restriction on the speech at issue be 'actually necessary' to achieve its interest."[527] Smolla reiterates that the Court strongly emphasized a clear articulation of the nature of harm resulting from the falsity: "[T]he plurality stated that '[t]here must be a direct causal link between the restriction imposed and the injury to be prevented.'"[528] In other words, traditional categories of legal harm can

---

518 The Court, considering the impact on creative expression, noted: "The Act by its plain terms applies to a false statement made at any time, in any place, to any person. It can be assumed that it would not apply to, say, a theatrical performance." *Id.* at 722 (citing Milkovich v. Lorain J. Co., 497 U.S. 1, 20 (1990)).

519 *Id.* ("Here the lie was made in a public meeting, but the statute would apply with equal force to personal, whispered conversations within a home.").

520 *Id.* at 723 ("And it does so entirely without regard to whether the lie was made for the purpose of material gain.").

521 *Id.* at 713.

522 *Id.* at 713–14, 726.

523 *Id.* at 729 ("[There was] no clear showing of the necessity of the statute, the necessity required by exacting scrutiny.").

524 *Id.* at 726 ("The Government points to no evidence to support its claim that the public's general perception of military awards is diluted by false claims such as those made by Alvarez.").

525 *Id.* at 725.

526 Rodney A. Smolla, *Categories, Tiers of Review, and the Roiling Sea of Free Speech Doctrine and Principle: A Methodological Critique of* United States v. Alvarez, 76 ALB. L. REV. 499, 514 (2013) (internal quotations omitted).

527 *Id.* at 514–15 (alteration in original) (internal citation and quotations omitted).

528 *Id.* at 515 (second alteration in original) (internal citation and quotations omitted).

offer a basis for falsity regulation. Further explained by Professor Martin Redish and Julio Pereyra, "[F]raud can justify government intervention, even after the Supreme Court's *Alvarez* decision."[529] The core inquiry, Professor Redish and Kyle Voils explain, is based "on the nature and severity of the harm, not on the commercial nature of the expression."[530] In a related vein, Professor Eugene Volokh has argued that

> being duped into hiring someone, or into opening your property to someone, based on affirmative lies would indeed count as a specific harm, even in the absence of physical property damage caused by the employee or visitor. . . .
>
> . . . .
>
> . . . And a public-spirited motive for getting a salary under false pretenses, or getting access to property under false pretenses, does not, I think, give a First Amendment immunity to the fraud.[531]

Finally, third, the Court examined the broader context of the falsity.[532] In particular, in analyzing the context of the prohibition on falsity, the Court found that the government's assertions of compelling interest[533] were not supported by its own prior acts. Specifically, the Court pointed out that rudimentary avenues of counterspeech[534] were

---

529 Martin H. Redish & Julio Pereyra, *Resolving the First Amendment's Civil War: Political Fraud and the Democratic Goals of Free Expression*, 62 ARIZ. L. REV. 451, 451 (2020).

530 Martin H. Redish & Kyle Voils, *False Commercial Speech and the First Amendment: Understanding the Implications of the Equivalency Principle*, 25 WM. & MARY BILL RTS. J. 765, 770 (2017).

531 Eugene Volokh, *Thoughts on the Court Decision Striking Down Idaho's "Ag-Gag" Law*, WASH. POST (Aug. 6, 2015), https://www.washingtonpost.com/news/volokh-conspiracy/wp/2015/08/06/thoughts-on-the-court-decision-striking-down-idahos-ag-gag-law/?utm_term=.c50182315390 [https://perma.cc/8YDA-AQAH].

532 United States v. Alvarez, 567 U.S. 709, 723 (2012) ("Permitting the government to decree this speech to be a criminal offense, whether shouted from the rooftops or made in a barely audible whisper, would endorse government authority to compile a list of subjects about which false statements are punishable. That governmental power has no clear limiting principle.").

533 *Id.* at 725 ("But to recite the Government's compelling interests is not to end the matter. The First Amendment requires that the Government's chosen restriction on the speech at issue be "actually necessary" to achieve its interest. There must be a direct causal link between the restriction imposed and the injury to be prevented. The link between the Government's interest in protecting the integrity of the military honors system and the Act's restriction on the false claims of liars like respondent has not been shown." (internal citations and quotations omitted)).

534 *Id.* at 726 ("The lack of a causal link between the Government's stated interest and the Act is not the only way in which the Act is not actually necessary to achieve the Government's stated interest. The Government has not shown, and cannot show, why counterspeech would not suffice to achieve its interest. The facts of this case indicate that the dynamics of free speech, of counterspeech, of refutation, can overcome the lie.").

not employed, in particular the creation of a website that would allow efficient confirmation of information, facilitating counterspeech.[535]

Thus, a close reading of *Alvarez* reveals that a constrained interpretation of the case is warranted. As succinctly stated by Professor Martin Redish, "[T]he First Amendment should have only a limited impact on regulation of false speech."[536] In summary, when properly drafted, falsity prohibitions can avoid duplicating the *Alvarez* pitfalls and potentially survive First Amendment scrutiny.

### 2. Content-Neutral Approaches

In addition to falsity prohibitions, however, other types of approaches unrelated to the falsity of particular content have also been proposed—additional disclosure and content labeling requirements, amplification restrictions, moderation liability, information reuse limitations, and personalization restrictions.[537] As one of us has already explained elsewhere,[538] content-neutral regulatory approaches to technology that focus on code functionality and information security harms—i.e., harms unrelated to the ideas presented by the content—can survive First Amendment scrutiny.[539]

#### a. Disclosure and Labeling Requirements

In lieu of outright bans on particular types of content, a fakery law might adopt a strategy of requiring additional clarifying disclosures—either through a separate clarifying filing or through a labeling requirement for the relevant content. Such disclosures would in many cases survive judicial scrutiny. Even in the sensitive context of mandatory disclosure and labeling requirements for political content, *Citizens United v. Federal Election Commission* signals that multiple

---

535 *Id.* at 729 ("There is, however, at least one less speech-restrictive means by which the Government could likely protect the integrity of the military awards system. A Government-created database could list Congressional Medal of Honor recipients. Were a database accessible through the Internet, it would be easy to verify and expose false claims.").

536 JAY B. STEPHENS & MARTIN H. REDISH, THE INTELLECTUAL GODFATHER OF COMMERCIAL SPEECH PROTECTION 5 (2017), https://s3.us-east-2.amazonaws.com/washlegal-uploads/upload/legalstudies/conversationswith/CWSummer2017.pdf [https://perma.cc/WJ3S-PUM4].

537 *See supra* notes 497–502 and accompanying text.

538 *See generally* Andrea M. Matwyshyn, *Hidden Engines of Destruction: The Reasonable Expectation of Code Safety and the Duty to Warn in Digital Products*, 62 FLA. L. REV. 109 (2010).

539 The hallmarks of successfully drafting such approaches are content neutrality, liability predicated on knowledge, and speaker-identity neutrality. *Id.* at 145–57.

types of disclosure and labeling approaches are viable.[540] In *Citizens United*, the Court expressly reaffirmed that disclaimers and disclosure requirements on political speech can survive First Amendment challenge in Internet contexts, finding "no constitutional impediment to the application of BCRA's disclaimer and disclosure requirements to a movie broadcast via video-on-demand" and stating that "there has been no showing that, as applied in this case, these requirements would impose a chill on speech or expression."[541] Thus, the Court specifically upheld disclosure requirements as applied to Citizens United's Internet communications.[542] Hence, even though they may burden the ability to speak,[543] disclosure requirements can survive scrutiny provided that they can be justified by a sufficient governmental interest. In other words, as explained by noted First Amendment scholars, even identity disclosure requirements[544] would likely pass First Amendment

---

540 558 U.S. 310, 312 (2010) ("[G]iven its complexity and the deference courts show to administrative determinations, a speaker wishing to avoid criminal liability threats and the heavy costs of defending against FEC enforcement must ask a governmental agency for prior permission to speak. The restrictions thus function as the equivalent of a prior restraint . . . .").

541 *Id.* at 371.

542 *Id.* at 319 ("The Government may regulate corporate political speech through disclaimer and disclosure requirements . . . ."). The Court explained that disclosure requirements on electioneering communications "'insure that the voters are fully informed' about the person or group who is speaking. . . . At the very least, the disclaimers avoid confusion by making clear that the ads are not funded by a candidate or political party." *Id.* at 368 (internal citation and quotations omitted). Similarly, citing prior precedent, the Court rejected Citizens United's challenge to a disclosure requirement:

> In *Buckley*, the Court upheld a disclosure requirement . . . . In *McConnell*, three Justices who would have found § 441b to be unconstitutional nonetheless voted to uphold BCRA's disclosure and disclaimer requirements. . . . And the Court has upheld registration and disclosure requirements on lobbyists, even though Congress has no power to ban lobbying itself. . . . For these reasons, we reject Citizens United's contention . . . .

*Id.* at 369 (internal citations omitted). Finally, the Court highlighted the critical role that disclosure requirements play:

> The First Amendment protects political speech; and disclosure permits citizens and shareholders to react to the speech of corporate entities in a proper way. This transparency enables the electorate to make informed decisions and give proper weight to different speakers and messages. . . . We find no constitutional impediment . . . .

*Id.* at 371.

543 *See id.* at 366–67 ("Disclaimer and disclosure requirements may burden the ability to speak, but they 'impose no ceiling on campaign-related activities' and 'do not prevent anyone from speaking'. . . . In *Buckley*, the Court explained that disclosure could be justified based on a governmental interest in 'providing[ing] the electorate with information' about the sources of election-related spending." (alteration in original) (internal citations and quotations omitted)).

544 An exception exists for situations where donors are exposed to likely retaliation and evidence of threats, harassment, or reprisals can be produced. *See id.* at 370.

scrutiny[545] in particular cases.[546] The primary concern with respect to their aggressive use is a historical and policy one: some types of labeling may hinder a key tool of furthering political discourse wielded by the Founding Generation—pseudonymous and anonymized speech.[547] In brief, in the context of Internet fakery regulation, some types of disclosure and labeling approaches are likely to pass First Amendment review.

### b. Amplification Restrictions

Another possible approach to limiting the impact of technology fakery involves restrictions on certain content-amplification conduct designed to create "noise." In Internet spaces, amplification conduct often intentionally drowns out particular content through amplifying other content, submerging the first from users' view and disrupting the public's quiet enjoyment of the Internet.[548] Indeed, at certain volumes of amplification, particularly when automated, the amplifying entity's behavior can often seem akin to the behavior of an attacker in a distributed denial of service attack. In other words, its character is more in line with conduct; content-neutral legal restrictions on amplification conduct do not necessarily offend the First Amendment.

In *Ward v. Rock Against Racism*, the Supreme Court explained that where the principal justification for governmental amplification guidelines arises from a desire to control noise levels in order to retain the character of a space and to avoid undue intrusion into areas of quiet enjoyment, the justification "satisfies the requirement that time, place, or manner regulations be content neutral."[549] The Court highlighted that the requirement of narrow tailoring is satisfied "so long as the . . . regulation promotes a substantial government interest that would be achieved less effectively absent the regulation" and "[s]o long as the means chosen are not substantially broader than necessary to achieve the government's interest. . . . 'The validity of [time, place, or manner] regulations does not turn on a judge's agreement with the responsible decisionmaker concerning the most appropriate method for promoting significant government interests.'"[550] To wit, limiting

---

545 *See* MARTIN H. REDISH, THE ADVERSARY FIRST AMENDMENT: FREE EXPRESSION AND THE FOUNDATIONS OF AMERICAN DEMOCRACY 163 (2013).

546 *E.g.*, McIntyre v. Ohio Elections Comm'n, 514 U.S. 334 (1995).

547 For a discussion of historical anonymous speech examples, see, for example, Benjamin Barr & Stephen R. Klein, *Publius Was Not a PAC: Reconciling Anonymous Political Speech, the First Amendment, and Campaign Finance Disclosure*, 14 WYO. L. REV. 253 (2014).

548 For a discussion of the quiet enjoyment of the Internet, see Andrea M. Matwyshyn, *Privacy, the Hacker Way*, 87 S. CAL. L. REV. 1, 48 (2013).

549 491 U.S. 781, 792 (1989).

550 *Id.* at 799–800 (quoting United States v. Albertini, 472 U.S. 675, 689 (1985) (first and fourth alteration in original)). Further, as the Court explained in *Holder v. Humanitarian Law Project*,

amplification conduct in particular technology contexts does not necessarily close a channel of communication; it can merely limit the volume to preserve quiet enjoyment of a particular space.[551]

While some legal scholars have characterized computational amplification tools such as bots as a "medium of speech" potentially offering "a new, unfolding form of expression,"[552] such technology exceptionalism warrants hefty skepticism.[553] Technology tools such as bots are more akin to remixing soundboards and sound trucks than to a writer's new novel. The basis for their regulation arises not from the content of what is being amplified but from the act of amplification itself—the intrusive conduct of using the amplifier in particular ways, regardless of message. As such, legal restrictions on (humans') conduct using bots and other technology amplification tools are likely to fall within the parameters of existing Supreme Court case law on amplification conduct. Ergo, just as some falsity prohibitions and disclosure/labeling requirements will survive First Amendment scrutiny, so too will properly crafted Internet amplification restrictions likely pass muster.

### c. Moderation Liability

At least two forms of moderation liability currently coexist—one driven by an Internet-first framing arising from the Communications Decency Act,[554] as qualified by the Supreme Court in *Reno v. ACLU*,[555] and a second, driven by traditional concepts of publishers' duty of care from pre-Internet times as articulated by the Court in *New York Times v. Sullivan*.[556] Scholarly opinions diverge on the law's application to technology contexts: Professor Cass Sunstein has argued that "*New York Times Co. v. Sullivan* badly overshot the mark and that it is ill-suited to the current era."[557] Meanwhile, Professor David Logan has

---

U.S. persons' speech can be not only limited in its volume but even entirely curtailed in particularly sensitive contexts, such as those related to national security. 561 U.S. 1 (2010).

551 *See Ward*, 491 U.S. at 792.

552 *See* Lamo & Calo, *supra* note 162, at 989.

553 For First Amendment purposes, rather than a "medium," a tool such as a bot might be more accurately described as a computational artifact—a mere amplification tool that, at best, assists humans with the creation of derivative works and recombinations. For a discussion of computational artifacts in the philosophy of technology sense, see, for example, *Computing as a Creative Activity and Computational Artifacts*, UNIV. OF R.I., https://computing-concepts.cs.uri.edu/wiki/Computing_As_A_Creative_Activity_and_Computational_Artifacts [https://perma.cc/MN8H-RL2T]; *cf. Artifact*, PCMAG, https://www.pcmag.com/encyclopedia/term/artifact [https://perma.cc/6NHN-V726] ("referring to software artifacts" in the engineering sense).

554 47 U.S.C. § 230.

555 521 U.S. 844 (1997).

556 376 U.S. 254 (1964).

557 Cass R. Sunstein, *Falsehoods and the First Amendment*, 33 HARV. J.L. & TECH. 387, 396 (2020).

argued that the *New York Times v. Sullivan* standard,[558] coupled with technology, has changed the nature of the public square.[559] Thus, argues Logan, the Court's approach should evolve in order to advance discussions regarding Internet speech and moderation.[560] The issues of moderation add a new dimension to the fakery inquiry, requiring their own in-depth analysis. They fall outside the scope of this Article and will be addressed in future work.[561]

### d. Information Reuse Restrictions

As explained in Section II.B, a portion of the Internet fakery dynamics are driven by information collection and reuse in targeted ways aimed at particular populations. As a consequence, one possible way to mitigate Internet fakery dynamics might be through information reuse restrictions. While such restrictions may seem initially problematic, upon closer examination of *Sorrell v. IMS Health Inc.*,[562] we find that the Supreme Court has offered guidance as to what type of information reuse restriction may survive First Amendment scrutiny. Notably, in *Sorrell*, the Supreme Court validated privacy interests as a legitimate state interest for First Amendment purposes.[563] As Professor Anupam Chander and Uyên P. Lê correctly point out, *Sorrell* does not mean "the "death of privacy."[564] *Sorrell* similarly does not mean the death of Internet fakery regulation. As one of us has argued elsewhere,[565] *Sorrell* does not pose a meaningful obstacle to (correctly drafted) information privacy statutes, provided they are neutral as to the commercial or noncommercial identity of the restricted entities—the fatal flaw of the statute at issue in *Sorrell*.[566]

### e. Personalization Restrictions

Thus, finally, and perhaps most surprisingly, the act of personalization of content itself can potentially be statutorily and

---

558 In particular, Professor Logan notes that "[b]efore *New York Times*, a defamatory statement was presumed to be false, which meant that the defendant had to prove its truth." David A. Logan, *Rescuing Our Democracy by Rethinking* New York Times Co. v. Sullivan, 81 OHIO ST. L.J. 759, 791 (2020).

559 *Id.* at 800.

560 *Id.* at 812.

561 A thorough First Amendment analysis of moderation liability will be considered at length in a subsequent paper entitled *Superspreaders*. *See* Matwyshyn, *supra* note 502 (on file with author).

562 564 U.S. 552 (2011).

563 *Id.* at 580 ("Privacy is a concept too integral to the person and a right too essential to freedom to allow its manipulation to support just those ideas the government prefers.").

564 Anupam Chander & Uyên P. Lê, *Free Speech*, 100 IOWA L. REV. 501, 522 (2015) ("Does *Sorrell* mean 'the death of privacy'? No, but it suggests serious limits to privacy law.").

565 For a discussion of *Sorrell v. IMS Health*, see Matwyshyn, *supra* note 548, at 13.

566 These issues will also be discussed in greater detail elsewhere. *See* Matwyshyn, *supra* note 502.

regulatorily restricted in some cases. These types of personalization and suitability requirements have already faced First Amendment challenge in *Lowe v. SEC*.[567] In *Lowe*, the Supreme Court began to set forth the contours of personalization limitations that pass constitutional muster. Specifically, the Court pointed to the intent of the creator of an Internet newsletter (that was not personalized), exempting him from registration requirements as an investment advisor on First Amendment grounds.[568] These insights from *Lowe*, particularly in combination with the recognition of privacy interests in *Sorrell*, signal that a statutory approach limiting personalization conduct in content creation may survive First Amendment scrutiny. Thus, personalization conduct restrictions may survive First Amendment scrutiny on both free speech and press freedom grounds where it is necessary to protect market integrity, for fraud prevention, or to further other legitimate state interests,[569] particularly where the dissemination arises from a statutorily created special status for the speaker/publisher. These key First Amendment insights inform the regulatory framework for Internet fakery next introduced in Part III.

## III. REGULATING UNTRUSTWORTHINESS: THE NICE FRAMEWORK

> *[C]ode is followed by commentary and commentary by revision, and thus the task is never done.*
>
> *—Justice Benjamin N. Cardozo*[570]

Part I introduced the problems of Internet MIST and two new dynamics—the arrival of the Internet long con and the risk of a PSYOP industrial complex. Reframing these challenges into a single concept, Part II introduced computer science and philosophical notions of trustworthiness and its inverse, untrustworthiness. It then presented the First Amendment constraints on any derivative legal framework of technology untrustworthiness. Informed by these concepts, this Part introduces one such legal framework for categorizing Internet fakery—the NICE framework. After introducing NICE, a set of example regulatory initiatives are set forth, together with their analysis using the NICE framework.

---

567 472 U.S. 181 (1985).

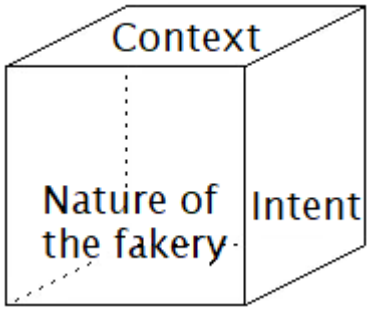
568 The Court noted that the newsletters "were offered to the general public on a regular schedule." *Id.* at 206. In other words, their publication was not tied to market events that impacted particular investors or that were tailored to investors' individual needs.

569 Some regulators have recently proposed various forms of personalization restrictions on Internet content, particularly in the context of mitigating election disinformation. *See, e.g.*, Ellen L. Weintraub, *Opinion: Don't Abolish Political Ads on Social Media. Stop Microtargeting.*, WASH. POST (Nov. 1, 2019, 6:51 PM), https://www.washingtonpost.com/opinions/2019/11/01/dont-abolish-political-ads-social-media-stop-microtargeting [https://perma.cc/AA2R-TZ63].

570 Benjamin N. Cardozo, *A Ministry of Justice*, 35 HARV. L. REV. 113, 117 (1921).

### A. *NICE and Precise: Nature, Intent/Knowledge, Context Evaluation*

The concept of untrustworthiness can be translated into a First Amendment–sensitive legal framework that consists of three separate inquiries or axes. The first axis involves an evaluation of the legal nature of the fake Internet content or conduct (N). The second axis in the evaluation involves the intent and knowledge of the faker and its legality (I). Finally, the third axis involves an assessment of the sensitivity of the context of the harm (C). These three elements are then blended into a single evaluation (E). Together, these four steps might be called *the NICE framework (NICE).*




#### 1. Axis 1: The Legal Nature of the Fakery

*To the rational mind nothing is inexplicable, only unexplained.*[571]

*—The Doctor*

Axis 1 of the NICE framework involves an evaluation of the legal nature of the untrustworthy technology content or conduct. For this, we turn for guidance to the philosophy scholarship on the nature of fakery, specifically work on lying and deception. While there is no universally accepted definition of lying[572] and deception,[573] some definitions are a better fit for a legal fakery framework than others. To wit, we borrow the spirit of a key insight from the work of Professor Thomas L. Carson[574]: the act of lying is an invitation to trust where the person making the offer knows or has reason to know that the trust is

571 *Doctor Who: The Robots of Death: Part One* (BBC One television broadcast Jan. 29, 1977).

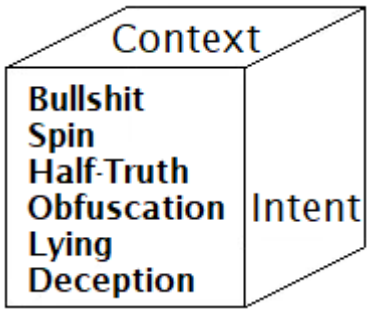
572 *The Definition of Lying and Deception*, STAN. ENCYCLOPEDIA OF PHIL. (Dec. 25, 2015), https://plato.stanford.edu/entries/lying-definition [https://perma.cc/9924-RNLW] ("There is no universally accepted definition of lying . . . .").

573 Definitions of deception often suffer from overbreadth of scope. *Id.* ("The principal problem is that [the definition of deception] is too broad in scope.").

574 THOMAS L. CARSON, LYING AND DECEPTION: THEORY AND PRACTICE (2010).

misplaced.[575] From this definitional insight springs a core distinction between lying and deception: Lying is extending the offer to trust and warranting its appropriateness[576]—i.e., making an untrustworthy offer. Deception occurs when that untrustworthy offer is accepted to the detriment of the acceptor[577]—i.e., successfully tricking a target with an untrustworthy offer. This key distinction bears echoes of the language of contract formation, a familiar paradigm for law that largely successfully coexists with First Amendment concerns.[578] In the remainder of this Section, we build out four additional categories that also translate into other legally cognizable categories of fakery.

We propose that the legal nature of all technology fakery can be categorized as falling into six conceptual categories of untrustworthy content: bullshit, spin, half-truth, obfuscation, lying, and deception (BSHOLD).[579] Each category of BSHOLD fakery will be discussed in turn below.




a. Bullshit

Philosopher Harry G. Frankfurt argues that "[o]ne of the most salient features of our culture is that there is so much bullshit."[580] In his seminal work, *On Bullshit*, he draws a distinction between lying and "bullshitting," arguing that lying constitutes a conscious act of deception whereas "bullshitting" is "indifference to how things really

575 *Id.* at 3.

576 *Id.*

577 *Id.*

578 *But see, e.g.*, Abigail Stephens, Note, *Contracting Away the First Amendment?: When Courts Should Intervene in Nondisclosure Agreements*, 28 WM. & MARY BILL RTS. J. 541 (2019).

579 As explained by Professor Thomas Carson, an untruth involves an invitation to trust where the person making the offer knows or has reason to know that the content is untrue. *See* CARSON, *supra* note 574, at 3. While philosophers will debate the meaning of truth, for our purposes the operative components of the differentiation involve the misplaced invitation to trust.

580 HARRY G. FRANKFURT, ON BULLSHIT 1 (2005).

are."[581] Frankfurt's framing[582] has been critiqued for both its failure to map cleanly to practical applications[583] and its failure to comfortably map to legal categories offered by First Amendment jurisprudence. As such, we explicitly reject Frankfurt's definition of bullshit and offer a different, First Amendment–sensitive legal replacement: *Bullshit refers to expressive fakery created for satirical or creative purposes that contains obvious exaggeration and embellishment in the eyes of a reasonable person.*[584]

For example, consider a music video available on YouTube where a Grammy-nominated[585] performer clad in Holstein-themed garb asserts many times sequentially, "I'm a cow."[586] No reasonable person is likely to believe that she is, in fact, a bovine, no matter how many times she says "Moo." This is prime cut, grade A Internet bullshit—it is creative, funny, engaging, and utterly (udderly?) ridiculous. It is also expression that a First Amendment analysis would protect. Thus, our definition of bullshit involves only obviously fake, creative, and satirical content that is uncontroversially protected by the First Amendment. This reframing avoids the ambiguities and pitfalls of Frankfurt's bullshit definition.

By way of a second example, consider the various animal face filters and lenses available for streaming applications and social media

---

581 *Id.* at 33–34. Jonathan Swift's description in 1710 of an English politician nails this phenomenon: "He never yet considered whether any Proposition were True or False, but whether it were convenient for the present Minute or Company to affirm or deny it." Jonathan Swift, *The Art of Political Lying*, EXAMINER, Nov. 9, 1710.

582 For a legal application of Frankfurt's framing see, for example, Rebecca Giblin, *Fat Horses & Starving Sparrows*, OVERLAND (2018), https://overland.org.au/previous-issues/issue-232/feature-fat-horses-starving-sparrows [https://perma.cc/MK59-LKFD].

583 For a critique of Frankfurt's case study selection as "strange and esoteric," see Robert Kane, *Response to Fischer, Pereboom, and Vargas*, *in* FOUR VIEWS ON FREE WILL 166, 166–83 (2007); *see also* Artem Kaznatcheev, *On Frankfurt's Truth and Bullshit*, THEORY, EVOLUTION & GAMES GRP. (May 4, 2019), https://egtheory.wordpress.com/2019/05/04/truth-and-bullshit [https://perma.cc/8W8L-WAT8].

584 For a related but different framing, see, for example, Catherine J. Ross, *Incredible Lies*, 89 U. COLO. L. REV. 377, 382 (2018) ("Bullshit, if recognizable as such, falls within the domain of expression I argue the law protects precisely because it is so out of bounds that no reasonable person would believe it.").

585 *Artist Doja Cat*, GRAMMY AWARDS, https://www.grammy.com/grammys/artists/doja-cat/287205 [https://perma.cc/FU9A-9BEF].

586 Doja Cat, *Doja Cat—"Mooo!" (Official Video)*, YOUTUBE (Aug. 10, 2018), https://www.youtube.com/watch?v=mXnJqYwebF8 (last visited Nov. 24, 2021) ("I'm a cow. . . . I'm a cow. I'm not a cat, I don't say meow.").

sites.[587] Cat filters in particular are a popular[588] overlay that falsifies videos by replacing human body parts with cat-looking body parts. The resulting videos are obviously doctored for comic effect, and they are unlikely to deceive any human viewers of sound mind. Let us continue discussing cat filters and turn to the next category—spin.

b. Spin

In a now–(in)famous Internet courtroom incident, a Texas defense attorney experienced a technical difficulty with a Zoom filter, resulting in his image being projected to the judge and opposing counsel as a white, distressed kitten. The lawyer/kitten plaintively announced, "I'm here live. I'm not a cat"[589] to the court, as he frantically fumbled with filter settings, trying to restore his visual humanity.[590] While the filter itself might be classified as an act of humor and bullshit, in this case, the lawyer's proclamation that he is, in fact, not a cat might be classified as something else—spin.

*Spin refers to the strategic framing of content*[591] *for advantage.*[592] Although the lawyer's internal image appeared to the naked eye to be a cat, his audio feed informed the judge that he was "here live."[593] In this way, he attempted to strategically reframe the unexpectedly hairy encounter. His statement attempted to channel the court's attention away from the talking white feline on the screen that belied his

---

587 The most commonly used Snapchat filter in 2017 was one that added cute pink furry ears. Alan Loughnane, *These Were the Most Used Snapchat Filters This Year*, JOE, https://www.joe.ie/tech/used-snapchat-filters-year-611236 [https://perma.cc/5BKZ-EDJK]; *see also, e.g.*, Lucas Matney, *Facebook Brings New Masks, Filters and Reactions to Messenger Video Chat*, TECHCRUNCH (June 26, 2017, 11:32 AM), https://techcrunch.com/2017/06/26/facebook-brings-filters-reactions-and-new-masks-to-messenger-video-chat [https://perma.cc/7M5A-FB66].

588 Many similar cat filters are available. *Cat Filter Vectors Images*, SHUTTERSTOCK, https://www.shutterstock.com/search/cat+filter+vectors [https://perma.cc/4FG6-8VF2].

589 Daniel Victor, *"I'm Not a Cat," Says Lawyer Having Zoom Difficulties*, N.Y. TIMES (May 6, 2021), https://www.nytimes.com/2021/02/09/style/cat-lawyer-zoom.html [https://perma.cc/SVZ3-95P8].

590 *Id.* This was not the first time that an unintentional use of cat filters in an inappropriate setting had made the news. *See, e.g.*, Agence France-Presse & Guillaume Lavallee, *Pakistan Politician Does Livestream with Cat Whiskers, Ears*, RAPPLER (June 15, 2019, 10:41 PM), https://www.rappler.com/world/regions/south-central-asia/233157-pakistan-politician-livestream-with-cat-filter-on [https://perma.cc/US7U-LBNA]; *Police Use Cat Filter in Murder Conference: When Live-Streaming Goes Wrong*, WEEK (July 23, 2019), https://www.theweek.co.uk/102393/police-use-cat-filter-in-murder-conference-when-live-streaming-goes-wrong [https://perma.cc/RJA5-L3LH].

591 For Thomas Carson, spin is putting an interpretation on undisputed events or facts. This interpretation will be biased, but not necessarily incorrect. CARSON, *supra* note 574, at 57–58.

592 The phenomenon of spin is related to the framing effect in behavioral psychology, which is that the way a decision problem is presented can influence the choice made by the decision maker. The metaphor is that of a visual presentation of the problem. Amos Tversky & Daniel Kahneman, *The Framing of Decisions and the Psychology of Choice*, 211 SCIENCE 453 (1981).

593 *See* Victor, *supra* note 589.

statement and toward the audio of his voice, highlighting his preparedness for the hearing.

The concept of "spinning" or presenting information in an optimally advantageous way for a speaker is certainly not new or legally unknown. Puffery is a well-established construct in both (pre-Internet) law of marketing[594] and securities regulation.[595] Or consider the functionality of photo-editing software. Editing images to make humans appear more in line with culturally specific standards of beauty is not a new practice. Indeed, an iconic image of President Lincoln is believed to have been a composite of Lincoln's head on Southern politician John Calhoun's body.[596] However, at some point, these acts of strategic framing and spinning may cross over into more problematic territory—half-truth,[597] lying,[598] and deception.[599]

### c. Half-truth

In technology contexts we often see fakery involving strategic omissions. For example, consider the infamous case of The Shed. In November 2017, the most highly rated of all the 18,000 restaurants in London listed on the travel information platform Tripadvisor[600] was one called The Shed at Dulwich.[601] Problematically, however, this dining "destination" was not actually a restaurant; it was a garden shed that had been hyped on Tripadvisor with fake reviews by friends and acquaintances of its Internet fakery-savvy owner.[602] Allegedly because of his prior success with writing fake reviews, the owner decided to see whether he could make a completely fake restaurant into a hit.[603] At first, he found excuses to turn potential customers away, but eventually, after reaching the number one spot, he invited three tables of customers to a meal at The Shed for a "press night," where he served them

---

594 *FTC Fact Sheet: It Looks Good . . . But Is It True?*, FTC, https://www.consumer.ftc.gov/sites/default/files/games/off-site/youarehere/pages/pdf/FTC-Ad-Marketing_Looks-Good.pdf [https://perma.cc/RZQ5-4GWM].

595 David F. Bandimere, Release No. 6521 (ALJ Mar. 27, 2019), https://www.sec.gov/alj/aljorders/2019/ap-6521.pdf [https://perma.cc/CW8R-HSY4].

596 Hany Farid, *Photo Tampering Throughout History*, https://www.cc.gatech.edu/~beki/cs4001/history.pdf [https://perma.cc/4GEP-3TX4].

597 *See infra* Section III.A.1.c.

598 *See infra* Section III.A.1.e.

599 *See infra* Section III.A.1.f.

600 TRIPADVISOR, https://www.tripadvisor.co.uk [https://perma.cc/7AE3-B25G].

601 Oobah Butler, *I Made My Shed the Top-Rated Restaurant on TripAdvisor*, VICE (Dec. 6, 2017, 6:44 AM), https://www.vice.com/en_us/article/434gqw/i-made-my-shed-the-top-rated-restaurant-on-tripadvisor [https://perma.cc/7GTC-9WKU].

602 *Id.*

603 *Id.*

microwave-ready meals.[604] Thus, The Shed was itself an Internet half-truth, a fictional restaurant driven by Internet fakery and hype that eventually served a real meal.[605]

*A half-truth involves a strategic blend of accurate description, material omission, and falsity that invites misplaced trust.* A construct familiar to law, strategic omissions are already considered in various bodies of law such as contract and securities regulation. When material,[606] they give rise to sanction through civil[607] and criminal means.[608] But the challenges of fake reviews such as those that propelled The Shed to culinary "standing" also lead us into our next category of fakery—obfuscation.

### d. Obfuscation

In a famous scene of *The Thomas Crown Affair*,[609] a business tycoon who recreationally steals art enters a crowded art museum to commit a heist, wearing a bowler hat and carrying a briefcase. A remote police team and insurance investigator vigilantly watch him through an intranet. As police move in to arrest him, he puts down the briefcase and picks up an identical, different one that is conveniently waiting for him. Suddenly, a swarm of haberdashed coconspirators arrive, walking in different directions, sometimes passing each other, swapping out identical-looking briefcases. The police team starts arresting bowler-hatted museum patrons to no avail—their target is gone. The swarm of bowler-hatted decoys with briefcases offers an example of fakery

---

604 Good Morning Britain, *The Fake Restaurant that Was London's Top Rated on Trip Advisor*, YOUTUBE, at 02:55 (Dec. 7, 2017), https://www.youtube.com/watch?v=yN_eTBe3NQ4 (last visited Nov. 24, 2021). The Shed at Dulwich is not the only fake restaurant to be ranked as the top spot to dine in a town as a result of fake positive reviews. Another example is Ristorante Scaletta, supposedly in the northern Italian town of Moniga del Garda, an imaginary restaurant created on Tripadvisor by several restaurant owners with support from the foodie publication Italia a Tavola. *See È il 1° in Classifica su TripAdvisor Ma il Ristorante di Moniga non Esiste*, ITALIA A TAVOLA (June 22, 2015, 2:32 PM), https://www.italiaatavola.net/cultura-media-lifestyle/stampa-web-tv-app/2015/6/22/e-il-1-in-classifica-su-tripadvisor-ma-il-ristorante-di-moniga-non-esiste/40173 [https://perma.cc/5KNB-AFZR].

605 Whether any legally actionable harm occurred in the case of The Shed likely depends on an assessment of the materiality of the half-truth and the nature of the relevant end user license agreements. But, for an example with more direct, quantifiable harm, consider a hiring or housing "pairing" algorithm that uses low-quality training data and consequently screens out all candidates of a particular race or gender for certain opportunities.

606 For a discussion of material omissions, see, for example, John C. Coffee, Jr., *Reforming the Securities Class Action: An Essay on Deterrence and Its Implementation*, 106 COLUM. L. REV. 1534, 1574 (2006).

607 For a discussion of contract omissions, see, for example, Susan Rogers Finneran, *Knowing Silence of Nonentrepreneurial Information Is Not Sporting*, 59 ALB. L. REV. 511, 522 (1995).

608 For a historical discussion of criminal culpability under securities laws for omission, see, for example, John G. Sobieski, *The Uniform Securities Act*, 12 STAN. L. REV. 103, 193–95 (1959).

609 THE THOMAS CROWN AFFAIR (United Artists, Irish DreamTime & Metro-Goldwyn-Mayer 1999).

through *obfuscation* or *noise*, hiding certain useful alternative information in a glut of extraneous information.

*Obfuscation or noise refers to the strategic creation of large amounts of content that makes material opposing information hard to find.* For example, in the context of a brutally adversarial litigation process, a common obfuscatory tactic of aggressive counsel in discovery involves overproduction of documentation in response to a production request in order to "bury" opposing counsel in paperwork.[610] This type of "document dump" forces opposing counsel to spend large amounts of time wading through nonresponsive information in order to find a needle in a haystack, a responsive document.[611]

In Internet contexts, recent hashtag takeovers by K-pop stans illustrate a somewhat parallel noise dynamic: K-pop fans have flooded Twitter with images of their favorite performers to dilute the visibility of content marked with a racist hashtag.[612] The behaviors themselves are in line with the mechanics of distributed denial of service attacks (DDoS).[613] In a DDoS[614] attack, an attacker attempts to make a website or application unavailable by sending it a large volume of traffic from multiple sources, often using a botnet made up of compromised machines or devices.[615] From the point of view of the website owner suffering a DDoS attack, the noise from the attack traffic makes requests for the site from ordinary legitimate users hard to identify;[616] from the point of view of ordinary users, the DDoS attack makes the content of

---

610 *See* Andrew J. Felser, *Document Production: Burying Responsive Documents Earns $10,000 Sanction*, A.B.A. (Jan. 31, 2017), https://www.americanbar.org/groups/litigation/committees/pretrial-practice-discovery/practice/2017/document-production-burying-responsive-documents-earns-10000-sanction (last visited Nov. 24, 2021).

611 In mergers and acquisitions, the parallel obfuscation involves burying liabilities that exist within the assets of the company in large quantities of due diligence materials.

612 Andrew Morse & Queenie Wong, *K-Pop Stans Take over Racist Hashtags on Twitter*, CNET (June 4, 2020, 9:44 AM), https://www.cnet.com/news/k-pop-stans-take-over-racist-hashtags-on-twitter-bts [https://perma.cc/8E9P-NCHU].

613 Christos Douligeris & Aikaterini Mitrokotsa, *DDoS Attacks and Defense Mechanisms: Classification and State-of-the-Art*, 44 COMPUT. NETWORKS 643 (2004).

614 DDoS is an act of computer intrusion under the Computer Fraud and Abuse Act (CFAA). For a historical discussion of the CFAA and DDoS, see, for example, Neal Kumar Katyal, *Criminal Law in Cyberspace*, 149 U. PA. L. REV. 1003 (2001).

615 *What Is a DDoS Attack?*, AKAMAI, https://www.akamai.com/uk/en/resources/ddos-attacks.jsp [https://perma.cc/JQC7-Q2F3].

616 For a short time during an attack in 2013, more traffic was sent to YouTube by bots than by humans. YouTube engineers hypothesized (perhaps as a joke) that this might cause "the Inversion," in which their fraud detection systems would label bot traffic as legitimate and legitimate traffic as fraud. Michael H. Keller, *The Flourishing Business of Fake YouTube Views*, N.Y. TIMES (Aug. 11, 2018), https://www.nytimes.com/interactive/2018/08/11/technology/youtube-fake-view-sellers.html [https://perma.cc/W2JK-KZ9J].

the website hard to discern, because it is hard to access or is unavailable. In the NICE framework, these are examples of obfuscation or noise.[617]

However, again, the specifics of the conduct in context frame the extent of harm. In some cases, the nature of the conduct is more than merely obscuring other information; the faker has warranted trustworthiness of content and context. When an explicit invitation to trust untrustworthy content is extended, that conduct qualifies as our next fake content category—lying.

### e. Lying

In 2015, the U.S. Environmental Protection Agency announced that some Volkswagen diesel vehicles[618] had been equipped with illegal defeat devices.[619] In other words, the cars used software in the vehicle's electronic control module (ECM) that sensed whether the vehicle was undergoing an emissions test, artificially altering the results to appear more favorable under test conditions.[620] During emissions tests the ECM produced compliant emissions, but in other circumstances the ECM switched to a "road calibration" scheme which reduced the effectiveness of emissions control.[621] The Volkswagen Group ultimately spent "tens of billions of dollars on regulatory fines and vehicle buybacks in the [United States] and [European Union],"[622] and executives faced criminal charges in both the United States and

---

617 In Thomas Carson's taxonomy, they are examples of concealing information. The difference between concealing information, withholding information, and deception is discussed in Section 2.III.3 of his book *Lying and Deception*. CARSON, *supra* note 574, at 56–57.

618 The problem was not confined to Volkswagen: other diesel manufacturers also came under suspicion. Gideon Lichfield, *Volkswagen's Emissions Scandal Is Just One Piece of a Larger Betrayal by Automakers*, QUARTZ (Apr. 25, 2016), https://qz.com/668553/volkswagens-emissions-scandal-is-just-one-piece-of-a-larger-betrayal-by-automakers [https://perma.cc/MQ75-CZUE].

619 A "defeat device" in a vehicle is defined by regulation as "an auxiliary emission control device (AECD) that reduces the effectiveness of the emission control system under conditions which may reasonably be expected to be encountered in normal vehicle operation and use." 40 C.F.R. § 86.1803-01 (2021).

620 The sensing was based on inputs including speed, steering wheel position, barometric pressure, and the duration of the engine's operation. Notice of Violation Letter from Phillip A. Brooks, Dir., Air Enf't Div., Off. of Civ. Enf't, to Volkswagen AG, Audi AG & Volkswagen Group of America, Inc. (Sept. 18, 2015), https://www.epa.gov/sites/production/files/2015-10/documents/vw-nov-caa-09-18-15.pdf [https://perma.cc/A4T8-6WPR]; *see also Mercedes-Benz Fined $1.5 Billion for Emissions Cheating*, CBS NEWS (Sept. 15, 2020, 6:54 AM), https://www.cbsnews.com/news/mercedes-benz-emissions-cheating-1-5-billion-fine-daimler [https://perma.cc/75RD-GFVD].

621 Brooks, *supra* note 620, at 4.

622 Megan Geuss, *German Regulator Says It Discovered New Illegal Software on Daimler Diesels*, ARS TECHNICA (June 24, 2019, 12:18 PM), https://arstechnica.com/cars/2019/06/german-regulator-says-it-discovered-new-illegal-software-on-daimler-diesels [https://perma.cc/6DGJ-6UQ8].

Germany.[623] In this scenario, Volkswagen was engaged in technology-enabled lying.

*Lying refers to the act of unilaterally going on the record with an untrustworthy statement*, inviting reliance upon it.[624] Through its defeat devices, Volkswagen lied to environmental regulators: it unilaterally warranted particular test emissions as trustworthy to accurately reflect compliance with environmental standards and invited reliance on that warranty. Similarly, the FTC's and the Consumer Financial Protection Bureau's (CFPB) technology enforcement is often predicated on untrustworthy assertions contained in companies' own advertising or website privacy policies—unilateral warranties that invite misplaced reliance.[625] Or, in a securities regulation context, a public company's false statement in its periodic disclosures that no material problem exists related to, for example, its information security[626] would constitute lying under our definition; it is conduct that potentially gives rise to a basis for both SEC enforcement[627] and investor litigation.[628] It reflects a situation where a company goes on the record with an assertion about its internal technology operations, knowing that even the absence of disclosure will be taken as a warranty of no material omission. Regardless of whether any particular investor navigates to the investor relations section of the corporate website and reads the 10K containing the material omission, by making the warranty, in the eyes of the SEC, the company invites misplaced reliance on untrustworthy content—i.e., carries out the act of lying. However, if particular investors rely on the lie to their detriment and take steps in reliance upon it, the lie crosses over into an act of deception.

---

623 Hiroko Tabuchi, Jack Ewing & Matt Apuzzo, *6 Volkswagen Executives Charged as Company Pleads Guilty in Emissions Case*, N.Y. TIMES (Jan. 11, 2017), https://www.nytimes.com/2017/01/11/business/volkswagen-diesel-vw-settlement-charges-criminal.html [https://perma.cc/2A8L-XSMX].

624 *See* CARSON, *supra* note 574, at 15–45.

625 For a discussion of FTC and CFPB unfairness authority, see, for example, Joshua D. Wright, *The Antitrust/Consumer Protection Paradox: Two Policies at War with Each Other*, 121 YALE L.J. 2216 (2012).

626 DIV. OF CORP. FIN., SEC, CF DISCLOSURE GUIDANCE: TOPIC NO. 2: CYBERSECURITY (Oct. 13, 2011), https://www.sec.gov/divisions/corpfin/guidance/cfguidance-topic2.htm [https://perma.cc/8T4Q-HD89].

627 *See, e.g.*, Press Release, SEC, SEC Charges Eight Companies for Failure to Disclose Complete Information on Form NT (Apr. 29, 2021), https://www.sec.gov/news/press-release/2021-76 [https://perma.cc/YU7G-WT8V].

628 Gregory A. Markel, Daphne Morduchowitz, Vincent A. Sama, Catherine B. Schumacher & John P. Hunt, *First Securities Class Action Complaint Filed in 2021 Following Disclosure of Cyberattack on SolarWinds Corporation*, SEYFARTH (Jan. 21, 2021), https://www.seyfarth.com/news-insights/first-securities-class-action-complaint-filed-in-2021-following-disclosure-of-cyberattack-on-solarwinds-corporation.html [https://perma.cc/HN3R-GHUP].

### f. Deception

Between 2013 and 2015, Facebook and Google were allegedly scammed out of over $120 million through an email scheme involving fraudulent invoices for allegedly purchased equipment.[629] The scam succeeded through phishing,[630] in which an attacker sends deceptive emails or other online messages that are faked to appear to come from a reputable source. If the victim is persuaded, the directed actions generally defraud the target for the benefit of the attacker and/or result in the victim's machine being infected with malware for later exploitation.[631] In other words, the attacker engaged in lying to the targets and tricked the targets into taking action to their detriment based on untrustworthy content or conduct. This is the final category of technology fakery—deception.

*Deception involves successfully tricking a target with lying, where the target acts in reliance on the untrustworthy content or conduct in a manner that causes detriment to the target.* While lying is a unilateral act of going on the record with fakery and inviting reliance, deception is the bilateral act of warranting trustworthiness of fakery that successfully tricks the target into taking action in reliance. As previously described, phishing attacks, in particular, rely on triggering misplaced trust.[632] They deceive through impersonation, attempting to convince the target of not only trustworthy content but also the trustworthiness of the sender, often engaging in conduct that is already prohibited, such as trademark infringement[633] and using sender information that

---

629 Shaun Nichols, *Five Years in the Clink for Super-Crook Who Scammed Google, Facebook Out of $120m with Fake Tech Invoices*, REGISTER (Dec. 20, 2019, 12:00 PM), https://www.theregister.co.uk/2019/12/20/facebook_google_hacker_five_years [https://perma.cc/Z9C3-P2N7].

630 Josh Fruhlinger, *What Is Phishing? How This Cyber Attack Works and How to Prevent It*, CSO (Sept. 4, 2020, 3:00 AM), https://www.csoonline.com/article/2117843/what-is-phishing-how-this-cyber-attack-works-and-how-to-prevent-it.html [https://perma.cc/7QPV-RU4U].

631 The perpetrator was eventually sentenced in Manhattan federal court to five years in prison. Press Release, Dep't of Just., Lithuanian Man Sentenced to 5 Years in Prison for Theft of over $120 Million in Fraudulent Business Email Compromise Scheme (Dec. 19, 2019), https://www.justice.gov/usao-sdny/pr/lithuanian-man-sentenced-5-years-prison-theft-over-120-million-fraudulent-business [https://perma.cc/M64X-TZTS].

632 Some phishing attacks use deceptive voicemail or SMS/text messages in the place of deceptive emails; these are sometimes known as vishing or smishing. David Bisson, *6 Common Phishing Attacks and How to Protect Against Them*, TRIPWIRE (Oct. 13, 2021), https://www.tripwire.com/state-of-security/security-awareness/6-common-phishing-attacks-and-how-to-protect-against-them [https://perma.cc/F793-DDNT].

633 For example, brands that have been frequently impersonated in phishing messages include major banks, PayPal, Microsoft, and Netflix. Adrien Gendre, *Phishers' Favorites: After Five Quarters, Microsoft Is Unseated by PayPal*, VADE (Nov. 7, 2019), https://www.vadesecure.com/en/phishers-favorites-q3-2019 [https://perma.cc/TXU3-YUW3].

resembles[634] that of the impersonated sender in order to intentionally cause confusion.[635]

But how do we analyze the less nefarious deceptions under the NICE framework? For example, how does a phishing attack differ from the situation where a friend merely copies and pastes the wrong link into an email? For this analysis we move to the second axis of the NICE framework—the faker's intent and knowledge.

### 2. Axis 2: Intent and Knowledge of the Faker

*Never gonna tell a lie and hurt you.*

*—Rick Astley*[636]

Rickrolling is an Internet prank[637] in which a faker deceives a target, tricking the target into clicking on a link that appears trustworthy but unexpectedly leads to a video of the 1987 pop hit "Never Gonna Give You Up" by Rick Astley.[638] In practical terms, a cheeky friend[639] tricking you with a Rickroll at worst triggers mild annoyance.[640] On a technical level, however, misrepresenting the content of the link in order to elicit a target's click constitutes a deception that is not dissimilar to a phishing attempt. Imagine if the linked Rick Astley video is not really a harmless act of Rickrolling from a friend; instead imagine the scenario where an attacker posing as your friend sends you a link to a malicious pretender video that impersonates

634 Fake domain names used for phishing may use homographs such as "micros0ft.com." Evgeniy Gabrilovich & Alex Gontmakher, *The Homograph Attack*, 45 COMMC'NS ASS'N FOR COMPUTING MACH. 128 (2002).

635 For a discussion of phishing, see, for example, Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 632–33 (2014) (discussing recent FTC phishing enforcement).

636 Rick Astley, *Rick Astley—Never Gonna Give You Up (Official Music Video)*, YOUTUBE, at 00:56 (Oct. 25, 2009), https://www.youtube.com/watch?v=dQw4w9WgXcQ (last visited Nov. 24, 2021).

637 The prank began on the 4chan site around May 2007, with a purported link to a preview for *Grand Theft Auto IV*. *Rickroll*, KNOW YOUR MEME, https://knowyourmeme.com/memes/rickroll [https://perma.cc/Y7GG-T4TC].

638 For an experiential introduction to definitely avoiding Rickrolling, see *How Not to Roll: An Experiential Introduction to Avoiding Rickrolling*, PENN STATE PILOT LAB, https://www.pilotlab2.org/hownottoroll [https://perma.cc/TA7K-PAKK].

639 The same is true of a cheeky footnote. *See id.*

640 James Tuckerman, *If You Thought Rickrolling Was Annoying, Check Out "Shredding,"* ANTHILL (Mar. 10, 2010), http://anthillonline.com/if-you-thought-rickrolling-was-annoying-check-out-shredding [https://perma.cc/6R2E-2AS7].

Astley's iconic one.[641] As you admire Astley's stylish trench coat,[642] malware infects your machine, and "lulz"[643] swiftly become replaced with "pwns."[644] Now consider a third scenario where your friend, a(n unironic) fan of Rick Astley's music, intends to send you a link to a news article about a local political event but instead accidentally copies and pastes the link from another window that is playing Rick Astley's music. As these examples demonstrate, although three acts of deception can initially technologically manifest in similar ways, the intent and knowledge behind each (and any subsequent harm)[645] can be importantly different. To wit, intent and knowledge comprise the second axis of a NICE analysis.

As previously discussed, the law is comfortable with assessments of intent and knowledge of defendants, and such analyses are present in various legal regimes to provide granularity in analysis.[646] For example, let us once again look to contract law, where deceptive intent and knowledge are derived partially from an analysis of the contractual relationship as a whole, including implicit power imbalances and equitable concerns.[647] In particular, contract law imposes implied duties of good faith in performance and breach mitigation, and as part of its analysis the court incorporates its objective assessment of the parties' intent.[648] Contract law also explicitly contemplates questions of fraud and misrepresentation through the lens of intent, dividing misrepresentation into intentional and innocent rubrics.[649] Materially different treatment is afforded by contract law to honest mistakes as

---

641 While using a Rick Astley video as a malware delivery vehicle may initially seem far-fetched, Rickrolling was indirectly implicated in the sale of NSA-linked exploits by a criminal consortium in 2016. Joseph Cox, *Someone Rickrolled the Bitcoin Auction for NSA Exploits*, VICE (Aug. 18, 2016, 5:20 AM), https://www.vice.com/en_us/article/yp3njm/someone-rickrolled-the-bitcoin-auction-for-nsa-exploits [https://perma.cc/FZT7-F4LK].

642 Ellie Kirwin, *Rick Astley Breaks Silence on Never Gonna Give You up Coat Theft "There Were Hundreds,"* EXPRESS (Sept. 30, 2020, 10:03 AM), https://www.express.co.uk/celebrity-news/1341761/Rick-Astley-never-gonna-give-you-up-trench-coat-stolen-video-news-latest-update [https://perma.cc/Q2DD-R2XV].

643 *Definition of "lulz,"* COLLINS DICTIONARY, https://www.collinsdictionary.com/us/dictionary/english/lulz [https://perma.cc/8HUZ-KNJ3].

644 *What Does "Pwn" Mean? And How Do You Say It?*, MERRIAM-WEBSTER, https://www.merriam-webster.com/words-at-play/pwn-what-it-means-and-how-you-say-it [https://perma.cc/U892-VKJC].

645 *See* discussion *infra* Section III.A.3.

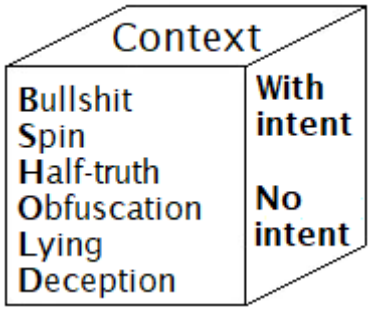
646 *See* discussion *supra* Section I.C.

647 *See, e.g.*, 1 EMP. COORDINATOR BENEFITS § 1:113 (2021) (discussing employer-employee power imbalances in contracting). For a historical discussion, see, for example, Arthur Allen Leff, *Unconscionability and the Code—the Emperor's New Clause*, 115 U. PA. L. REV. 485 (1967).

648 *See, e.g.*, 2 CORP. COUNS.'S GUIDE TO DISTRIB. COUNSELING § 17:19 (2021). For a historical discussion, see, for example, Steven J. Burton, *Breach of Contract and the Common Law Duty to Perform in Good Faith*, 94 HARV. L. REV. 369 (1980).

649 *See, e.g.*, 17A C.J.S. CONTS. § 230 (2021). For a historical discussion, see, for example, Clare Dalton, *An Essay in the Deconstruction of Contract Doctrine*, 94 YALE L.J. 997, 1039 (1985).

opposed to intentional misstatements or omissions.[650] A similarly nuanced analysis of intent and knowledge can be brought to bear in analysis of technology fakery through the NICE framework.




Returning to Volkswagen's defeat devices as an example of lying, as the plea agreement in *United States v. Volkswagen* articulates, the defeat devices in question were developed and deployed in Volkswagen's vehicles with the knowledge and supervision of Volkswagen employees[651]—strong evidence of intent to engage in illegal conduct. As another example of intent and knowledge playing a qualifying role, consider a purchase of fake followers for a social media account. The purchaser obfuscates the true popularity of the account, and such a purchase is an intentional act. However, some researchers have suggested the possibility that a sudden large increase in the number of fake followers of a particular politician's account may have been the result of a non-supporter purchasing them to embarrass the politician without the knowledge of the politician.[652] If so, in such cases the obfuscation was not deliberate on the part of the account owner. Thus, intent and knowledge are central to a nuanced determination of appropriate sanction for Internet fakery. So too is the sensitivity of the context in which the fakery causes harm.

---

650 For a historical discussion, see, for example, Note, *Equitable Relief for Unilateral Mistake of Fact*, 30 HARV. L. REV. 637 (1917).

651 Plea Agreement, United States v. IAV GmbH, No. 16-CR-20394 (E.D. Mich. Dec. 18, 2018); *see also* Martin Gelter, *Employee Participation in Corporate Governance and Corporate Social Responsibility* 25 (Eur. Corp. Governance Inst., Working Paper No. 322, 2016).

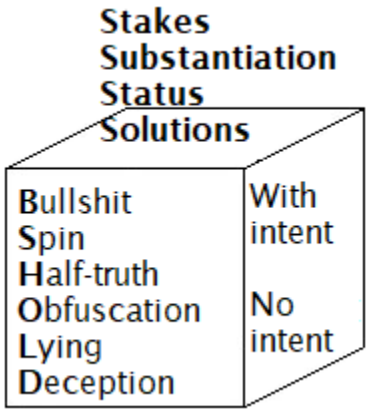
652 Axel Bruns, Darryl Woodford & Troy Sadkowsky, *Towards a Methodology for Examining Twitter Follower Accession*, FIRST MONDAY (Apr. 2014), https://firstmonday.org/article/view/5211/3864 [https://perma.cc/5R72-7JXE].

### 3. Axis 3: Context Sensitivity

> *We didn't focus on how you could wreck this system intentionally [when designing the Internet].*
>
> —*Vinton Cerf*[653]

In his seminal book *Code and Other Laws of Cyberspace*, Professor Lawrence Lessig recounts a story of two neighbors, Martha and Dank, and an incident involving the unfortunate poisoning of Dank's dog. Lessig highlights that "[o]ne difference was the nature of the space, or context, where their argument was happening"[654]: it was within a massively multiple online game, and the dog was a virtual object in the game. As a result, compared with a dog poisoning in an offline context, the stakes are arguably lower, or at least different, as are the possible solutions.[655] Yet, the nature of Martha and Dank's relationship as (virtual) neighbors was a complicating element of the context.[656] A fact-intensive inquiry is unavoidable.

The third axis of the NICE framework recognizes this insight and involves the (legally cognizable) sensitivity of the context in which the technology fakery occurs. Just as particular acts of fakery can vary in intent and knowledge, so too the sensitivity of the context varies, affecting the ultimate impact of the fakery. A context sensitivity analysis can be divided into four separate prongs of inquiry—stakes, substantiation, status, and solutions.




The first prong, the stakes of the context, involves an assessment of the severity of possible (legally cognizable) harms that could arise for

653 *See* Rainie & Anderson, *supra* note 20.

654 LESSIG, *supra* note 196, at 11.

655 *Id.* at 13–14.

656 *Id.* at 10–14.

targets of the fakery, as well as the number of targets impacted. For example, in a situation where Internet fakery may have directly or contributorily caused death, the consideration of stakes would push the assessment toward a more negative outcome for the faker and a strong connection to a set of grievous existing legal harms—homicide/manslaughter and wrongful death. Consider the situation where an attacker spoofs the identity of an employee with administrative privileges to log in and change settings on a hospital system maintaining ventilators, resulting in the death of patients. Or consider the scenario where the ill-fated Fyre Festival might have resulted in deaths due to lack of potable water; in such a circumstance, the victims would have presumably relied on the untrustworthy hype about luxury accommodations and assumed that potable water was available on the island.[657] If the context is sensitive enough to potentially result in or contribute to the death of unwitting targets, the context sensitivity is high and potentially merits greater sanction for the fakery.

The second prong, substantiation, asks whether a shared baseline assessment method on point exists, and whether the fakery at issue violates the expectations of the target on the point of that shared baseline.[658] Through first establishing the existence of a shared baseline and valuation method,[659] courts and legislators can eliminate the need for complex legal or philosophical discussions of "truth." This approach of explicitly judicially incorporating shared baselines is not novel: contract law has long recognized the value of incorporating externally created baselines to resolve disputes between parties. For example, contract law cases regularly include judicial analysis in reliance on external pricing mechanisms,[660] the norms of the course of dealings in the parties' industry,[661] word meaning as determined by outside experts,[662] and other similar objective baseline determination methods to resolve disputes where contracts are silent or unclear. Similarly, as the FTC has recognized, the image of a lock triggers a shared consumer expectation of certain levels of encryption and security in information

---

657 *See* discussion *supra* Section I.A.1.b.

658 This question of shared baselines will be explored in greater detail in *Superspreaders*. *See* Matwyshyn, *supra* note 502 (on file with author).

659 In general, shared baselines fall into four categories based on their method of creation: 1. Hierarchy, 2. Expertise, 3. Legacy, and 4. Process. *See* Matwyshyn, *supra* note 502 (on file with author).

660 For a discussion of the use of open price terms in contracts, see, for example, Mark P. Gergen, *The Use of Open Terms in Contract*, 92 COLUM. L. REV. 997 (1992).

661 For a discussion of course of dealing between merchants, see, for example, Lisa Bernstein, *Merchant Law in a Merchant Court: Rethinking the Code's Search for Immanent Business Norms*, 144 U. PA. L. REV. 1765, 1781–82 (1996).

662 *See, e.g.*, Frigaliment Importing Co. v. B.N.S. Int'l Sales Corp., 190 F. Supp. 116 (S.D.N.Y. 1960) (discussing various governmental standards around "chickens").

transmission on a website.[663] If a website presents users the lock icon without also meeting the implicitly understood level of security, a shared baseline developed through both expertise and process is violated. Therefore, the context sensitivity is greater and the fakery arguably more worthy of sanction.

Third, the status prong involves an examination of the types of power imbalances disfavored by various bodies of law including contract law and criminal law—e.g., forms of severe information imbalance,[664] manipulation,[665] or access limitations to external assistance.[666] The analysis of this prong should align with, for example, contract law notions of equitable and relational[667] concerns, including the prior course of dealings of the parties, the involvement of third parties either in reliance or to further the subterfuge, as well as the extent of reliance by the target of the fakery. Just as in information-imbalanced contract situations, ambiguities should be construed in favor of the target subjectively under this prong. Thus, this prong injects some subjective analysis into an otherwise objective inquiry. In other words, while the second prong of substantiation engages with the objective question of deviation from recognized baselines, the third prong of status allows for recognition of subjective experiences of the fakery target, i.e., further recognizing the particularities of each context.

The final prong examines possible solutions. Just as in contract law, where financial compensation and other methods of correction can sufficiently remediate the harm, the sensitivity of the context (and the severity of the harm) are generally deemed less severe. The extent of accurate prior threat modeling and effectiveness of risk mitigation enter

---

663 Ashkan Soltani, *FTC.gov Is Now HTTPS by Default*, FTC (Mar. 6, 2015, 11:00 AM), https://www.ftc.gov/news-events/blogs/techftc/2015/03/ftcgov-now-https-default [https://perma.cc/D2RS-ADTB].

664 Concerns over severe information imbalance are visible in traditional contract doctrines such as unconscionability with its focus on unfair surprise and oppression, as well as duress and coercion. For a historical discussion of duress and coercion in contract, see, for example, Margaret Jane Radin, *Market-Inalienability*, 100 HARV. L. REV. 1849 (1987).

665 Manipulation is a concern visible in bodies of law such as securities regulation, where it is defined as engaging in conduct that "creat[es] a false or misleading appearance of active trading in any security . . . or a false or misleading appearance with respect to the market for any such security" or using or employing "any manipulative or deceptive device or contrivance." Securities Exchange Act of 1934, 15 U.S.C. §§ 78i(a)(1), 717c-1. As described in the Act, the rules "are designed to prevent fraudulent and manipulative acts and practices, to promote just and equitable principles of trade . . . and, in general, to protect investors and the public interest." *Id.* § 78o–3(b)(6). For a historical discussion of securities regulation and market manipulation, see, for example, *Market Manipulation and the Securities Exchange Act*, 46 YALE L.J. 624 (1937). Manipulation also raises dignitary concerns over autonomy and respect. *See, e.g.*, Helen Norton, *The Government's Lies and the Constitution*, 91 IND. L.J. 73, 79 (2015).

666 Depraved heart murder, for example, considers access limitations to assistance. For a historical discussion of depraved heart murder, see, for example, 40 C.J.S. HOMICIDE § 41 (2021).

667 For a discussion of relational concerns and contract, see, for example, Ian R. MacNeil, *Relational Contract: What We Do and Do Not Know*, 1985 WIS. L. REV. 483.

the analysis as well. The most severe situations involve harms for which solutions are not readily available through damages awards or through orders of specific performance—loss of life, harms where time was of the essence, national security harms, and the like. In such fakery cases, criminal sanctions might be favored over merely civil sanctions. As the nature of harm becomes more complex and difficult to cleanly articulate, this signal should serve as a harbinger of the lack of compensability and the failure of damages awards as adequate remedy. For instance, psychological and electoral harms intentionally inflicted on civilian populations with malicious intent through the use of PSYOP are an example of a severe harm where precise measurement is challenging. As the Army PSYOP Field Manual explains, even when results of PSYOP are visible, precise quantification and causal ties are difficult to prove.[668] This does not change the reality of the negative impact and the appropriateness of sanction as a national security harm.

This four-prong context sensitivity analysis enables the creation of a common language around both economic and nonpecuniary harms across different categories of technology fakery. In particular, it avoids narrow Internet exceptionalist analysis that would hinder harmonization across international boundaries. Similarly, its technology-neutral framing assists with identification of relevant traditional legal approaches that can be married into analyses of technology harms. In this way, it is more likely to successfully avoid unintended spillover effects that damage established bodies of law—a problem already visible in some technology regulatory contexts.[669]

### 4. The Evaluation

The evaluation process in the NICE framework involves three steps set forth in the table that follows. The first step involves a categorization of the legal nature of the fakery. Second, a determination of a faker's level of malicious intent and knowledge occurs. Third, the evaluation determines the legally cognizable sensitivity of the context and its impact on the harm resulting from the fakery. Finally, appropriate legislative and regulatory sanction is then considered based on the results of this evaluation, engaging with traditional legal paradigms as much as possible. The more obviously malicious the intention, the greater the intent or level of knowledge of likely harm arising from the fakery, and the more sensitive the context, the more appropriate the imposition of legal sanction.

---

668 *See* HEADQUARTERS, *supra* note 254, at 11–14.

669 For a discussion of how technology exceptionalist paradigms of computer intrusion are damaging contract law, see Matwyshyn, *supra* note 385.

| **NATURE OF FAKERY** | **INTENT AND KNOWLEDGE**: INTENTIONAL/KNOWLEDGE <==> UNINTENTIONAL/NO KNOWLEDGE | | |
|---|---|---|---|
| **BULLSHIT** | Humor content on the internet | Intentional cat filter | Unintentional cat filter |
| **SPIN** | Directed advertising toward partners<br>*Theft/Fraud; Contract; Tort, e.g., tortious interference with contract* | Photoshop image correction<br>Twitter "puffery" about a particular company's value<br>*Depends* | Corrected eye direction in Facetime |
| **HALF-TRUTH/ OMISSION** | Price discrimination/demonetization by algorithm<br>*Regulatory, Criminal, Civil Sanction* | <= Algo error=><br>*Depends* | (Non-manipulated) Info in a crowd-sourced recommendation system |
| **OBSTRUCTION/ NOISE** | DDoS<br>*Intrusion/Vandalism; Crim & Civil Sanction* | Fake followers<br>*Depends* | Traffic overload of a website |
| **LYING** | Defeat device<br>Tech copyright infringement<br>*Regulator, Criminal, Civil Sanction* | <= Deep fakes =><br><= Buggy code =><br>*Depends* | Reposting fake news believing it true |
| **DECEPTION** | Phishing; MiTM<br>*Fraud: Regulatory/Criminal & Civil Sanction* | Rick-rolling<br>*Depends* | Copy/pasting wrong link into an email to a friend |

LEVEL OF SANCTION

CONTEXT SENSITIVITY

To engage with a concrete application, let us return to cat filters. Clearly bullshit under our categorization, the application of cat filters by a user is usually intentional; however, as mentioned in Part III, there have been cases where it has also occurred unintentionally. Cat filters, even when unintentional and inappropriate in context, are unlikely to cause any legally cognizable harm if a product of user error. However, consider the scenario where the cat filter is applied by a third party through an act of computer intrusion. Here the context of the fakery modifies the analysis and the appropriate sanction: the sanction results not from the content itself but from the conduct underlying the Internet fakery—the intrusion. Thus, the appropriate target for any regulatory intervention is not the involuntary cat filter user, it is instead the intentional perpetrator of the crime of computer intrusion that resulted in the nonconsensual cat filter.

Next consider a doctored video[670] where facial features have been adjusted,[671] or where objects or people have been edited out[672]—a practice with a particularly problematic history.[673] Editing of images may also occur unexpectedly in some technology contexts—without the knowledge of the person producing the images. For example, some applications correct eye-gaze direction,[674] which although benign, has been described by some users as "creepy."[675] Alternatively, text giving important context for an image may be removed, for example, from a newspaper clip circulated online,[676] or a video's speed may be changed at critical moments, giving a false impression.[677] In these scenarios, defamation claims may offer one viable remedy or could evolve to offer one.

Now consider a different, particularly challenging half-truth scenario—one where a flawed machine learning system is relying on inadequate training data to generate classifications of participants in, for example, a government benefits program, and impacted users might experience financial, safety, and dignitary harms.[678] Here, while the legally problematic nature of the fake content is straightforward, the analysis of intent/knowledge and context becomes more complicated.

---

670 *Photoshop*, ADOBE, https://www.adobe.com/products/photoshop.html [https://perma.cc/7AUZ-XSQW].

671 Photoshop Training Channel, *How to Use the Face-Aware Liquify in Photoshop*, YOUTUBE (June 21, 2016), https://www.youtube.com/watch?v=2zhgvNfJTnM (last visited Nov. 24, 2021).

672 *Adobe Creates Artificial Intelligence Removing Entire Objects from Video Clips*, HITECHER (Apr. 5, 2019), https://hitecher.com/news/adobe-creates-artificial-intelligence-removing-entire-objects-from-video-clips [https://perma.cc/6KKD-DWTZ].

673 *See Former Soviet Union Circa 1930*, BRONX DOCUMENTARY CTR. ALTERED IMAGES, http://www.alteredimagesbdc.org/stalin [https://perma.cc/R2MA-DXX3].

674 Andrew O'Hara, *Hands on with Apple's FaceTime Attention Correction Feature in iOS 13*, APPLEINSIDER (July 3, 2019), https://appleinsider.com/articles/19/07/03/hands-on-with-apples-facetime-attention-correction-feature-in-ios-13 [https://perma.cc/C7AJ-JE2E].

675 Mark's Tech, *iOS 13 Beta 3 FaceTime Attention Correction is CREEPY . . . But Cool*, YOUTUBE (July 7, 2019), https://www.youtube.com/watch?v=vGD7ayC9vnI (last visited Nov. 24, 2021).

676 For example, an Internet image that appeared to be Mahatma Ghandi dancing with a woman in an evening gown had been edited to remove the original newspaper caption that the "Ghandi" was an impersonator. *Misleading & Photoshopped Images in Social Media*, GURUPRASAD'S PORTAL, http://guruprasad.net/posts/misleading-photoshopped-images-in-social-media [https://perma.cc/GY3Q-CKJP].

677 Megan Geuss, *AP: Video Expert Says White House Clip of CNN Reporter Was Likely Doctored [Updated]*, ARS TECHNICA (Nov. 9, 2018, 9:45 AM), https://arstechnica.com/gaming/2018/11/ap-video-expert-says-white-house-clip-of-cnn-reporter-was-likely-doctored [https://perma.cc/K4Q5-PS2P].

678 Michele Gilman, *AI Algorithms Intended to Root Out Welfare Fraud Often End Up Punishing the Poor Instead*, CONVERSATION (Feb. 14, 2020, 8:45 AM), https://theconversation.com/ai-algorithms-intended-to-root-out-welfare-fraud-often-end-up-punishing-the-poor-instead-131625 [https://perma.cc/7VD8-25NB].

Finally, an Internet pump-and-dump scheme offers a straightforward example of an intentional deception in a sensitive context.[679] Untrustworthy content and perhaps artificial amplification conduct induce reliance to the financial detriment of targets in such a scenario. The content is intended to defraud, and the context sensitivity is already legally recognized as high by securities regulation—investors suffer from an information imbalance and the government interest in preserving financial market stability through trustworthy financial disclosures is well established.

### B. *A NICE Future for Fakery Regulation: Addressing MIST*

The preceding Sections of this Article have introduced the NICE framework for addressing technology fakery. This concluding Section crystalizes some of the possible legislative and regulatory lessons of the framework. It also highlights examples of specific types of regulatory interventions that may hold promise in line with the NICE framework.

#### 1. Addressing Manipulation

In addressing manipulation, as explained previously, a prohibition on false content remains a viable approach if carefully crafted. In summary, even after *Alvarez*, legal restrictions on false speech are most likely to survive First Amendment scrutiny where they reflect four criteria. First, they target a traditional category of harm previously recognized by the Court[680] unrelated to a particular (competing) communicative message.[681] In order for any outright prohibition on false content to survive First Amendment scrutiny, the framing of the prohibition must start from the identification of a narrow, specific traditionally recognized state interest. That list is short: preservation of fair bargaining in the marketplace, national security, public safety, administration of justice, and preservation of other core governmental functions. Second, the regulation is drafted in a manner that demonstrates a causal link between the restricted false content and the harm, implicitly underpinned by malicious intent/knowledge of the

679 For an example of recent SEC pump-and-dump enforcement, see Press Release, SEC, SEC Obtains Asset Freeze in Microcap Pump and Dump Scheme Targeting Elderly Retail Investors, (July 18, 2019), https://www.sec.gov/news/press-release/2019-136 [https://perma.cc/X8E8-WUW2].

680 Redish and Voils in particular highlight "five categorical harms to which false speech may conceivably give rise: (1) financial; (2) political; (3) reputational; (4) health and safety; and (5) interpersonal." Redish & Voils, *supra* note 530, at 794.

681 *See* Nat'l Inst. of Fam. & Life Advocs. v. Becerra, 138 S. Ct. 2361 (2018) (striking down a statute requiring notification of state-sponsored abortion services).

faker.[682] Third, the selected regulatory framing should be one that promises greater efficacy, explaining why less burdensome ways driven by counterspeech have failed or would prove ineffective.[683] Finally, the commercial versus noncommercial identity of the speaker should not be a determinative element under the statute.[684] If properly drafted, restrictions on falsity in particular technology contexts that meet these criteria can offer one possible regulatory intervention for addressing a portion of technology fakery.[685] As the standards for commercial and noncommercial speech continue to merge, so too does the treatment of natural versus corporate persons under the law. The legislative and regulatory approaches most likely to survive First Amendment scrutiny are those blind not only to the message of a speaker but also blind to both the content's commercial or noncommercial nature and the identity of the speaker as a human or corporate person.

As a first cut, however, the most promising approaches in addressing manipulation in the short term are those that rely on labeling and additional disclosure; they can be scaled quickly. For example, because First Amendment rights apply only to U.S. persons,[686] labeling requirements for content produced outside the United States by non-U.S. persons that target U.S. audiences would likely survive First Amendment scrutiny, particularly if limited to sensitive contexts such as election communications where foreign direct involvement is already legally regulated.[687]

### 2. Addressing Impersonation

Legal approaches that focus on prohibiting impersonation of identity of legal persons hold promise.[688] For example, they might target misidentification of source and user confusion in a manner reminiscent

---

682 United States v. Alvarez, 567 U.S. 709, 732 (2012) (Breyer, J., concurring).

683 *Id.* at 726–27.

684 Sorrell v. IMS Health Inc., 564 U.S. 552 (2011); *see, e.g.*, STEPHENS & REDISH, *supra* note 536, at 5 (discussing the "*strong presumption against distinguishing between commercial and non-commercial speech*" in particular when "the sole basis of distinction is that one speaker is commercial and the other speaker is non-commercial").

685 Professor Richard Hasen offers a cautionary note regarding the idea of creating a government arbiter of truthful information: "Truth commissions pose risks as well as harms, and it may be that counterspeech is the best we can do consistent with the First Amendment and the risks of the alternatives." Richard L. Hasen, *A Constitutional Right to Lie in Campaigns and Elections?*, 74 MONT. L. REV. 53, 57 (2013).

686 But see cases where political viewpoints were used as a basis for exclusion, deportation, and refusal of naturalization. *E.g.*, United States *ex rel.* Turner v. Williams, 194 U.S. 279 (1904); Harisiades v. Shaughnessy, 342 U.S. 580 (1952); Kleindienst v. Mandel, 408 U.S. 753 (1972).

687 *Cf.* 52 U.S.C. § 30121 (contribution and expenditure ban applied to "foreign national[s]").

688 Parody falls outside the scope of actionable impersonation for purposes of the NICE framework. *See supra* Section III.A.1.

of frameworks in trademark. While preservation of pseudonymity holds value in furthering discourse, impersonation of another existing person's identity (in a nonsatirical manner) is not an act of pseudonymity—it is an act of goodwill usurpation and source misidentification. For example, statutes that prohibit Internet use of another person's identity for purposes of fraud or voter manipulation might offer a natural extension of existing law: voter fraud and identity theft statutory frameworks on point such as the Identity Theft Penalty Enhancement Act[689] and similar statutes already exist in both state and federal law.[690] Additionally, state data security or privacy statutes might be expanded to limit the repurposing of particular categories of residents' personal information for any commercial or noncommercial purpose, creating private rights of action. Finally, defamation statutes might be updated with technology-neutral language, eliminating a portion of the current obstacles to the creation of more robust and equitable doctrines of Internet defamation.[691]

### 3. Addressing Sequestration

To mitigate sequestration, a statutory disclosure requirement might mandate the creation of a public repository where all variants of candidate-sponsored and PAC-sponsored political advertising targeting U.S. citizens must be filed shortly after first use. Such a repository could be jointly managed by the FEC, DOJ, and Cybersecurity and Infrastructure Security Agency and would serve national security and election integrity interests. It would offer both third-party researchers and government enforcers additional tools to identify foreign election interference and other problematic Internet fakery dynamics. Such an approach would also likely survive First Amendment scrutiny.[692] Another possible regulatory avenue involves strengthening disclosure requirements for paid endorsements. In particular, relevant regulators might strengthen existing disclosure requirements for social media influencers and increase enforcement actions targeting unlabeled paid promotions by social media

689 Identity Theft Penalty Enhancement Act, Pub. L. No. 108-275, 118 Stat. 831 (2004).

690 *See, e.g.*, Identity Theft Enforcement and Restitution Act, Pub. L. No. 110-326, § 202, 122 Stat. 3560 (2008).

691 For example, clarifying and potentially expanding defamation per se, offering new categories of equitable remedies, as well as expanding the law of injurious falsehood, may offer particular avenue for state approaches. In particular, states should consider expanding the availability of anti-SLAPP statutes.

692 However, where the speaker can demonstrate concerns of retaliation based on the message in the content, identity disclosure obligations may not survive First Amendment scrutiny. This limitation does not impact the ability to submit the content itself, however. *See* discussion *supra* Section II.B.2.a.

influencers.[693] Other approaches are also possible, including potentially an influencer or targeted advertising "broker dealer" registration and disclosure scheme modeled on securities regulation.[694]

### 4. Addressing Toxicity

In addressing toxicity, separate from any restriction on untrustworthy content, restrictions on certain types of amplification conduct and personalization/targeting may be appropriate. Such restrictions might be based on a broader set of governmental interests, including preserving quiet enjoyment of a space, preserving election security, or preserving market fairness.[695] Just as we restrict DDoS attacks[696] and bot-facilitated purchases of concert tickets,[697] artificial "straw amplification" of content created by U.S. persons may be restricted, recognizing that such a regulation should be cautious to avoid any assessment of the underlying message itself or a restriction of the underlying channel for unamplified use.[698] Similarly, Congress might instruct that agencies such as the FEC, FTC, SEC, and CFPB introduce targeted/personalized messaging prohibitions where critical

693 For an example of current social media influencer disclosure guidelines, see, for example, FTC, *supra* note 132.

694 This expansion would involve granting the FTC enhanced penalty authority and budgetary capacity to increase enforcement, as well as creating a parallel expansion of the FEC's approach to paid political endorsement disclosures and enforcement. One of us has previously advocated for the creation of an FTC Technology Practices Group that would work across agency lines to address cross-cutting technology fakery issues. *See* Matwyshyn, *supra* note 178.

695 Code amplification tools are not persons and do not possess First Amendment interests. The persons who wield them may hold First Amendment interests, but those interests are not unlimited. Creators of those tools can be regulated in line with current interpretations of the Commerce Clause.

696 *See* 18 U.S.C. § 1030; Press Release, Dep't of Just., Man Receives Maximum Sentence for DDoS Attack on Legal News Aggregator (June 11, 2020), https://www.justice.gov/usao-ndtx/pr/man-receives-maximum-sentence-ddos-attack-legal-news-aggregator [https://perma.cc/78MY-NUC8].

697 15 U.S.C. § 45c; Lesley Fair, *FTC's First BOTS Act Cases: Just the Ticket to Help Protect Consumers from Ticket Bots*, FTC (Jan. 22, 2021, 12:09 PM), https://www.ftc.gov/news-events/blogs/business-blog/2021/01/ftcs-first-bots-act-cases-just-ticket-help-protect-consumers [https://perma.cc/44NY-AKCD].

698 Compensated "straw amplification" restrictions might also fall under enhanced labeling requirements. In the spirit of "straw purchaser" restrictions on alcohol, tobacco, and firearms regulation, as used here, "straw amplification" refers to the idea of engaging paid amplifiers to pretend to be engaging in content creation on their own behalf—a service that would likely fall under Congress's Commerce Clause authority.

infrastructure integrity interests are at stake.[699] Additionally, registration and conduct reporting requirements for public relations firms in a manner parallel to the structures in place for various participants in financial markets might begin to mitigate the problems of fakery amplification by dark PR and PSYOP professionals for hire.[700]

## CONCLUSION

*Oh, what a tangled web we weave*

*When first we practise to deceive!*[701]

For five years, a ten-foot bronze spider sculpture created by artist Louise Bourgeois sat on Pier 14 of the Embarcadero in San Francisco, welcoming visitors and residents alike.[702] On loan from a private collector, the giant spider had originally been slated to remain on display for only eight months.[703] However, because of public fascination with the art, the run was extended.[704] An inspirational and somewhat controversial piece of art, the crouching spider triggered years of awe, appreciation, and debate.[705] Bourgeois's giant spiders offer an arachnid contrast to the tarantulas that lived in AT&T's basement and to Lucian's giant spiders. Unlike those menacing arachnids, Bourgeois's spiders engaged with the public without guile or risk of harm, sometimes causing spirited disagreements about their merits.

The model of Internet fakery regulation that offers us a path forward is one that leads us closer to the spirited debates caused by Bourgeois's spiders and further away from looming debacles of AT&T's and Lucian's spiders. As this Article has explained, the first step in this

---

699 Regulations motivated by concerns over election integrity, market integrity, and fair bargaining concerns will potentially survive First Amendment scrutiny, particularly if such restrictions are constructed within the scope of a broader registration framework such as the SEC's approach.

700 Tighter regulation of conduct in the dark PR industry as a sector of the economy would survive Commerce Clause analysis. For example, a registration regime based on securities regulation may be appropriate. For a discussion of the existing registration requirements for participants in financial markets, see, for example, David A. Lipton, *A Primer on Broker-Dealer Registration*, 36 CATH. U. L. REV. 899 (1987).

701 WALTER SCOTT, MARMION: A TALE OF FLODDEN FIELD, canto 6, stanza XVII (1808).

702 John Coté & Heather Knight, *"Crouching Spider" Saying Farewell to S.F.*, SFGATE (Feb. 8, 2012, 9:59 PM), https://www.sfgate.com/bayarea/article/Crouching-Spider-saying-farewell-to-S-F-3164801.php [https://perma.cc/65L2-E3H6].

703 *Id.*

704 *Id.*

705 *Id.* Other versions of Bourgeois's spiders remained in a different context—the San Francisco Museum of Modern Art. *Louise Bourgeois Spiders*, SFMOMA, https://www.sfmoma.org/exhibition/louise-bourgeois-spiders [https://perma.cc/DUH2-WG8N].

enterprise involves crafting a common baseline among policymakers and regulators around a legal concept of "untrustworthiness" through the NICE framework—an evaluation driven by the legal nature of fake content, the intent and knowledge of the faker, and the sensitivity of the context in which the fakery occurs.