University of Michigan Journal of Law Reform

Volume 54

2021

Publish, Share, Re-Tweet, and Repeat

Michal Lavi
Hadar Jabotinsky Center for Interdisciplinary Research of Financial Markets

https://doi.org/10.36646/mjlr.54.2.publish

PUBLISH, SHARE, RE-TWEET, AND REPEAT

Michal Lavi*

ABSTRACT

New technologies allow users to communicate ideas to a broad audience easily and quickly, affecting the way ideas are interpreted and their credibility. Each and every social network user can simply click "share" or "retweet" and automatically republish an existing post and expose a new message to a wide audience. The dissemination of ideas can raise public awareness about important issues and bring about social, political, and economic change.

Yet, digital sharing also provides vast opportunities to spread false rumors, defamation, and Fake News stories at the thoughtless click of a button. The spreading of falsehoods can severely harm the reputation of victims, erode democracy, and infringe on the public interest. Holding the original publisher accountable and collecting damages from him offers very limited redress since the harmful expression can continue to spread. How should the law respond to this phenomenon and who should be held accountable?

Drawing on multidisciplinary social science scholarship from network theory and cognitive psychology, this Article describes how falsehoods spread on social networks, the different motivations to disseminate them, the gravity of the harm they can inflict, and the likelihood of correcting false information once it has been distributed in this setting. This Article will also describe the top-down influence of social media platform intermediaries, and how it enhances dissemination by exploiting users' cognitive biases and creating social cues that encourage users to share information. Understanding how falsehoods spread is a first step towards providing a framework for meeting this challenge.

The Article argues that it is high time to rethink intermediary duties and obligations regarding the dissemination of falsehoods. It examines a new perspective for mitigating the harm caused by the dissemination of falsehood. The Article advocates harnessing social network intermediaries to meet the challenge of dissemination from the stage of platform design. It proposes innovative solutions for mitigating careless, irresponsible sharing of false rumors.

The first solution focuses on a platform's accountability for influencing user decision-making processes. "Nudges" can discourage users from thoughtless sharing of falsehoods and promote accountability ex ante. *The second solution*

---

* Ph.D. (Law). Research Fellow, Hadar Jabotinsky Center for Interdisciplinary Research of Financial Markets, Crisis and Technology; Postdoctoral Fellow, University of Haifa, Faculty of Law; Cyberlaw Fellow, Federmann Center Hebrew University; Cheshin Fellow, Hebrew University, Faculty of Law, 2018. I thank Michal Shur-Ofry and Emily Cooper. Special thanks are due to Nick Adkins, Hannah Basalone, Nicole Frazer, Sumner Truax, and their colleagues on the *University of Michigan Journal of Law Reform* staff.

I dedicate this Article to the memory of my mother—Aviva Lavi—who died suddenly and unexpectedly. My mother taught me to love knowledge and gave me the strength to pursue it. She will always be loved, remembered, and dearly missed.

focuses on allowing effective ex post facto *removal of falsehoods, defamation, and fake news stories from all profiles and locations where they have spread. Shaping user choices and designing platforms is value laden, reflecting the platform's particular set of preferences, and should not be taken for granted. Therefore, this Article proposes ways to incentivize intermediaries to adopt these solutions and mitigate the harm generated by the spreading of falsehoods. Finally, the Article addresses the limitations of the proposed solutions yet still concludes that they are more effective than current legal practices.*

TABLE OF CONTENTS

INTRODUCTION

> *In 2015, a man named Ariel Ronis came across a defamatory post on Facebook accusing him of racism. The post was not restricted to the publisher's friends but was public to all Facebook users. Over 6,000 individuals disseminated the post using the "share" button, without knowing whether the statements made about this person were actually true. Recipients of the post continued to share it while some even added comments condemning Ronis. The post went viral as it spread rapidly and garnered media attention. The man felt his good reputation was ruined and ended up committing suicide.*[1]
>
> *In the 2016 U.S. presidential election cycle, falsehoods and fake news were spread about both candidates. For example, it was rumored that Hillary Clinton helped to fund and arm ISIS,*[2] *that she and her campaign chief were running a pedophilia ring from the basement of a pizza parlor,*[3] *and that the Pope had endorsed Donald Trump,*[4] *even though none of this had actually happened. Many believe that these rumors and others like them influenced the results of the elections.*

The advent of technology and social media have revolutionized interpersonal communication. Within seconds, a message or a post can travel around the world and be viewed by thousands of users.[5] Individuals publish and spread messages thoughtlessly and almost automatically at the click of a button (e.g., publish, share, re-tweet). Sharing information has many benefits. Each and every internet user can replicate valuable ideas and raise public awareness of important issues, even though they are not affiliated with a press organization. As a result, free speech has become easily accessible

---

1. This is the story of Ariel Ronis, an official from Israel's Population, Immigration and Border Crossing Authority, part of the Ministry of Interior, who shot himself to death after a Facebook post accusing him of racism went viral. *See Interior Ministry Official Commits Suicide After Accusation of Racism Goes Viral*, JERUSALEM POST (May 24, 2015, 2:50 AM), https://www.jpost.com/israel-news/interior-ministry-official-commits-suicide-after-accusation-of-racism-goes-viral-403924 [hereinafter RONIS].

2. *See* YOCHAI BENKLER, ROBERT FARIS & HAL ROBERTS, NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS 139–44 (2018) ("The 'Hillary helped fund and arm ISIS' story depends on a rich shared narrative created by media that have longer and deeper purchase on the minds of those who are exposed to it.").

3. *See* BENKLER ET AL., *supra* note 2, at 3.

4. *Id.* at 142.

5. *See* Haewook Kwak, Changhyun Lee, Hosung Park & Sue Moon, *What Is Twitter, a Social Network or a News Media?*, 19 INT'L CONF. ON WORLD WIDE WEB 591, 591 (2010) (finding that every re-tweeted tweet on Twitter will reach an average audience of 1,000 people regardless of the number of the original publisher's followers).

to all. This brave new technological world has enabled individuals to voice their opinions on important social, political, and economic issues.

Yet, there is a flip side. Information wants to be free, but so does misinformation. The information revolution allows for the vast spreading of rumors, speculations, and assumptions about private individuals and public figures without any checks on the accuracy of the content. False rumors, defamation, and fake news stories are amazingly powerful and dangerous; it is difficult to reverse them;[6] they can have serious consequences for a person's reputation, and they can even cost life.[7] Falsehoods can also cause harm to their audiences and to society in general.[8] In rare cases, false stories can even result in physical harm to the *recipients* of the rumor. For example, fake news regarding potential cures for the Covid-19 pandemic resulted in people consuming "miracle cure[s]" that caused physical damage.[9] Such "misinformation pose[s a] threat to public health."[10]

Moreover, as false rumors spill into the digital ecosystem, they pollute the flow of information.[11] Consequently, it becomes more and more difficult to distinguish between true and false information and to engage in truthful discussions on matters of public importance. Thus, politics, democracy, and the public interest in general are impaired.[12]

Spreading falsehoods raises complex challenges as the more times people are exposed to falsehoods, the more credible they

---

6. *See* Neil Levy, *The Bad News About Fake News*, 6 SOC. EPISTEMOLOGY REV. & REPLY COLLECTIVE 20, 20 (2017) ("[F]ake news is more pernicious than most of us realise, leaving long lasting traces on our beliefs and our behavior even when we consume it know [sic] it is fake or when the information it contains is corrected.").

7. *See, e.g.*, RONIS, *supra* note 1.

8. *See* Cass R. Sunstein, *Falsehoods and the First Amendment*, 33 HARV. J.L. & TECH. 387, 388 (2020) ("Some falsehoods are harmful. They ruin lives. They lead people to take unnecessary risks or fail to protect themselves against serious dangers.").

9. Hugo Mercier, Opinion, *Fake News in the Time of Coronavirus: How Big Is the Threat?*, THE GUARDIAN (Mar. 30, 2020, 10:00 AM), https://www.theguardian.com/commentisfree/2020/mar/30/fake-news-coronavirus-false-information (describing a man exposed to fake news on the cure for Covid-19 who "ingest[ed] a product meant to clean fish tanks, as it contains chloroquine, a drug currently being tested (inconclusively so far) as a treatment for Covid-19" and died as a result).

10. *See COVID-19 Global Roundup: Conspiracy and Fake News Challenge Public Health and Big Tech*, CGTN (May 13, 2020, 4:27 PM), https://news.cgtn.com/news/2020-05-13/COVID-19-Global-Roundup-Conspiracy-and-fake-news-a-test-for-big-tech-QsqUXgDpHa/index.html.

11. *See* Omri Ben Shahar, *Data Pollution*, 11 J. LEGAL ANALYSIS 104, 105, 112–13 (2019) (treating "fake news" as "data pollution" that disrupts social institutions and public interests in a similar manner to environmental pollution).

12. *See* Richard L. Hasen, *Deep Fakes, Bots, and Siloed Justices: American Election Law in a Post-Truth World*, 64 ST. LOUIS U. L.J. 535, 540 (2020); Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 YALE J.L. & TECH. 106, 144–45 (2019); Anthony J. Gaughan, *Illiberal Democracy: The Toxic Mix of Fake News, Hyperpolarization, and Partisan Election Administration*, 12 DUKE J. CONST. L. & PUB. POL'Y 57, 68 (2017).

become.[13] These falsehoods start to feel so true that people believe them even when provided with evidence of their falsity.[14] Research reveals that falsehoods "diffused significantly farther, faster, deeper, and more broadly" than truthful content,[15] and in many cases, private efforts to refute a falsehood by publishing the truth not only fail to cancel out the falsehood's impact but can even increase its credibility.[16] Even when attempts to correct a falsehood do succeed in mitigating its influence, their effect remains limited, as they are often less widely viewed than the original falsehood.[17]

Victims of defamation can file suit when they have access to the alleged defamer's name. In their terms of service, many social networks such as Facebook and Linkedin require users to provide their real names.[18] Thus, victims of defamation can file a lawsuit against the publisher.[19] Filing an action against the original speaker and collecting damages, however, provides very limited redress because it cannot counteract complications arising from the wide-

---

13. Gerd Gigerenzer, *External Validity of Laboratory Experiments: The Frequency-Validity Relationship*, 97 AM. J. PSYCH. 185, 185, 192–93 (1984) ("[M]ere repetition of plausible but unfamiliar assertations increases the belief in the validity of the assertations, independent of their actual truth or falsity.").

14. Whitney Phillips, *The Toxins We Carry*, COLUM. JOURNALISM REV. (Fall 2019), https://www.cjr.org/special_report/truth-pollution-disinformation.php ("It shows that when people are repeatedly exposed to false statements, those statements start to feel true, even when they are countered with evidence. In short, a fact check is no match for a repeated lie.").

15. Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, 359 SCI. MAG. 1146, 1147 (2018).

16. *See* Gordon Pennycook, Tyrone D. Cannon & David G. Rand, *Prior Exposure Increases Perceived Accuracy of Fake News*, 147 J. EXPERIMENTAL PSYCH. 1865 (2018). Research also shows that false information might psychologically cancel out the influence of truthful statements. *See* Sander van der Linden, Anthony Leiserowitz, Seth Rosenthal & Edward Maibach, *Inoculating the Public Against Misinformation About Climate Change*, 1 GLOB. CHALLENGES 1, 2, 5 (2017).

17. *See* Mark Scott & Melissa Eddy, *Europe Combats a New Foe of Political Stability: Fake News*, N.Y. TIMES (Feb. 20, 2017), nyti.ms/2SEP0ej (discussing the unsuccessful trials of EU teams to correct fake news due to their dissemination).

18. *See Terms of Service 3.1*, FACEBOOK, https://www.facebook.com/terms.php [https://perma.cc/PE52-AGWJ] (last visited Nov. 2, 2020) ("[U]se the same name that you use in real life."); *User Agreement 2.1(2)*, LINKEDIN (Aug. 11, 2020), https://www.linkedin.com/legal/user-agreement [https://perma.cc/M2TS-EWAY] ("[Y]ou will only have one LinkedIn account, which must be in your real name."); Tal Z. Zarsky & Norberto Nuno Gomes de Andrade, *Regulating Electronic Identity Intermediaries: The "Soft eID" Conundrum*, 74 OHIO ST. L.J. 1335, 1336–37 (2013) (referring to the possibility of removing or blocking profiles that do not reflect a real identity); *This Woman Changed Her Name Just so She Could Log In to Facebook*, TIME (July 13, 2015, 1:17 AM), https://time.com/3955056/facebook-social-media-jemma-rogers-uk/ [https://perma.cc/SH23-EYJ6] (after creating an alias for her Facebook account, a woman resorted to legally changing her name to avoid being locked out of her account).

19. *See, e.g.*, Boulger v. Woods, 917 F.3d 471 (6th Cir. 2019) (dismissing a defamation case because the defendant added question marks to his allegedly defamatory tweets); *see also* Patrick H. Hunt, Comment, *Tortious Tweets: A Practical Guide to Applying Traditional Defamation Law to Twibel Claims*, 73 LA. L. REV. 559 (2013) (reviewing defamation lawsuits regarding libel claims for statements posted on Twitter, referred to as "twibel claims").

spread dissemination of falsehoods. In this context, § 230 of the Communications Decency Act (CDA) reflects the "internet exceptionalism" that "diverge[s] from regulatory precedents in other media."[20] It directs that "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."[21] Courts have interpreted § 230 broadly as providing immunity to internet users who share information of other users who are not providers.[22] Consequently, lawsuits against disseminators of posts authored by third parties are usually blocked. Moreover, filing an acaction against all users who share a falsehood is impractical due to their large number and the administrative costs of filing an action against each and every one of them. It might also be unfair to hold an individual accountable for content he did not originally author because, unlike conventional news outlets that have a duty to devote time and resources to vetting stories prior to publication,[23] a private citizen cannot be expected to ascertain the credibility of such content let alone have a duty to do so once content has already been published online.[24] Imposing liability on disseminators may also deter individuals from social sharing of content and have a chilling effect on free speech.[25] It can also dilute the liability of the person who created and published a post and, therefore, re-

---

20. Eric Goldman, *The Third Wave of Internet Exceptionalism*, TECH. & MKTG. L. BLOG (Mar. 11, 2009), https://bit.ly/2KGhOkP; *see also* John Perry Barlow, *A Declaration of the Independence of Cyberspace*, ELEC. FRONTIER FOUND. (Feb. 8, 1996), https://www.eff.org/cyberspace-independence; JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 77–78 (2019).

21. 47 U.S.C. § 230. For further discussion of this provision, see *infra* Section II.A.

22. *See* KOSSEFF, *supra* note 20, at 145–47; *see also, e.g.*, Barrett v. Rosenthal, 146 P.3d 510, 529 (Cal. 2006) ("Nor is there any basis for concluding that Congress intended to treat service providers and users differently."). It should be noted however that if the defendant authored content that accompanies republished content, he might become a provider of content and § 230 might not apply. *See* La Liberte v. Reid, 966 F.3d 79 (2d Cir. 2020); Loeb & Loeb LLP, La Liberte v. Reid, LEXOLOGY (July 15, 2020), https://www.lexology.com/library/detail.aspx?g=2e200267-569e-41aa-b7eb-f04939246c8b ("Reid herself had authored the content that accompanied the photograph of La Liberte and did not merely republish the photograph from another 'information content provider.' ").

23. *See* Vanessa S. Browne-Barbour, *Losing Their License to Libel: Revisiting § 230 Immunity*, 30 BERKELEY TECH. L.J. 1505, 1511–12 (2015) (exploring traditional standards of liability for defamation and explaining that traditional publishers face strict liability).

24. *See* Matt C. Sanchez, Note, *The Web Difference: A Legal and Normative Rationale Against Liability for Online Reproduction of Third-Party Defamatory Content*, 22 HARV. J.L. & TECH. 301, 311, 319–20 (2008) (explaining that unlike traditional media, "online speakers as a class do not have the experience or resources" to verify "the facts contained in every piece of information they reproduce," and, therefore, they should not bear liability).

25. For more in the related context of intermediaries (websites or public pages) that republish users' content and make it more visible, see Michal Lavi, *Taking Out of Context*, 31 HARV. J.L. & TECH. 145, 179–80 (2017).

duce his incentives to prevent harm.[26] It would also have administrative costs that exceed its benefits.[27]

Due to these considerations, this Article does not focus on the liability of users who disseminate third-party falsehoods. Instead, it focuses on online intermediaries, such as Facebook and Twitter, that design network tools and facilitate social sharing of organic content for profit.[28]

This Article further focuses on the extensive dissemination of negative false rumors through social media. The discussion is not limited to fake news about politicians and public figures but extends to lies and defamation about ordinary people. It focuses on social networks, such as Facebook and Twitter, due to their centrality and the socio-technological context, which facilitates sharing valuable content easily while also exacerbating speech-related harm.

The liability and accountability of intermediaries for user content currently exists in the realm of policy discussion; the law does not yet provide solutions to the problem of spreading falsehoods.[29] This Article aspires to bridge the gap by proposing to harness intermediaries to the mission of mitigating the dissemination of falsehoods. It proposes to promote accountability from the stage of platform design. It aims to offer an *ex ante* solution for accommodating the challenge of spreading falsehoods at the stage of the user's decision to post the falsehood and share it. It also proposes an *ex post* solution for mitigating the harm of falsehoods that have already been shared. Keeping this goal in mind, the Article is divided into the following parts:

Part I describes how ideas spread online. Drawing on network theory, psychology, marketing, and information systems, it outlines

---

26. Ronen Perry, *The Law and Economics of Online Republication*, 106 IOWA L. REV. (forthcoming 2021) (manuscript at 32–36), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3552301# (presenting the theory of dilution of liability and expanding it to liability of republishers).

27. *Id.* at 38 (addressing the administrative costs of legal actions against republishers).

28. *See* Kyunghee Lee, Byungtae Lee & Wonseok Oh, *Thumbs Up, Sales Up? The Contingent Effect of Facebook Likes on Sales Performance in Social Commerce*, 32 J. MGMT. INFO. SYS. 109, 110–11, 139 (2015) (explaining that social sharing enhances the flow of information, attracts users, and can be translated into sales). The Article addresses social sharing of organic content and not targeting of advertisements by intermediaries. For expansion on ad targeting of political advertisements for profit, see Editorial, *Twitter Is Banning Political Ads. If Facebook Won't, It Must at Least Moderate Them*, WASH. POST (Nov. 1, 2019), https://www.washingtonpost.com/opinions/twitter-is-banning-political-ads-if-facebook-wont-it-must-at-least-moderate-them/2019/11/01/9d3457c0-fc01-11e9-ac8c-8eced29ca6ef_story.html [https://perma.cc/GE93-5CVF].

29. The Article will address the current legal realm, which focuses on removal of harmful content *ex post*, allowing content to spread up until and even subsequent to removal. *See generally* James Grimmelmann, *The Platform Is the Message*, 2 GEO. L. TECH. REV. 217, 224–25 (2018).

the myriad of motivations for spreading falsehoods. Afterwards, it explores "bottom-up" social dynamics among users and "top-down" influences of intermediaries on social network platforms. These dynamics increase the likelihood that falsehoods will spread rapidly, inflicting severe harm on the defamed individual and the public's interest.

Part II explores the law governing the secondary liability of intermediaries. It argues that with respect to spreading falsehoods, current policy models fall short. Therefore, complementary mechanisms should be formulated. The Article will argue that due to the growing power and influence of intermediaries on the flow of information, they should bear more responsibility. As a first step, intermediaries should be accountable for embedding complementary mechanisms to accommodate for the damages of replication and dissemination of falsehoods.

Part III proposes solutions to mitigate the problem of spreading falsehoods. It argues that intermediaries, which encourage users to share content and profit from sharing, should promote user accountability. First, it suggests using "*nudges*" to influence user decisions to share content.[30] Nudges are expected to influence the context of user decision-making before publishing and sharing content and have the potential to mitigate the spread of falsehoods *ex ante*. The second solution focuses on efficient removal of harmful content *ex post facto*. Accordingly, an intermediary that designs features for simplifying dissemination should provide technology for efficient removal of the content shared from all profiles and locations. This solution is already applied by some social network platforms.[31] Choice architecture, however, "is value-laden, and reflects a particular set of preferences that should not be taken for granted."[32] Thus, the Article proposes incentives for intermediaries to adopt architecture for efficient removal and addresses their limitations. The proposals focus on the initial stage of platform design that have a much better chance of hindering the spread of false-

---

30. *See* RICHARD THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH AND HAPPINESS 6 (2009) (defining a nudge as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives").

31. For example, Facebook facilitates social sharing by embedding posts. Thus, if the original post is deleted, the content becomes unavailable in the profiles it was embedded in. *See Embedded Posts*, FACEBOOK, https://developers.facebook.com/docs/plugins/embedded-posts [https://perma.cc/TDV3-QB77] (last visited Sept. 19, 2020); Toby Headdon, *An Epilogue to Swenson: The Same Old New Public and the Worms that Didn't Turn*, 9 J. INTELL. PROP. L. & PRAC. 662, 666 (2014) (explaining that embedded links allows efficient removal from YouTube).

32. Michal Lavi, *The Good, the Bad, and the Ugly Behavio*r, 40 CARDOZO L. REV. 2597, 2671 (2019).

hoods. These proposals could have been particularly useful during the 2020 election cycle, as they mitigate the harm of defamation and fake news stories to the reputation of individuals, including candidates for president. By allowing intermediaries to correct failures in the marketplace of ideas, the proposals promote people's sense of reality, keeping them more invested in facts and real news. Thus, the proposals also preserve the public interest.

I. SHARING CONTENT ON SOCIAL NETWORKS

Why do falsehoods spread within social networks? Bringing together multidisciplinary insights, this part explores the process of dissemination online, explains the motivation behind spreading falsehoods, and provides answers to this initial question. Afterwards, it explores the "bottom-up" social dynamics among users and the "top-down" influence of intermediaries on social dissemination. It concludes that these dynamics can exacerbate the harm caused by falsehoods.

A. *Why Do False Rumors Spread?*

Falsehoods, defamation, and fake news spread rapidly and gain credibility. For example, in the 2016 U.S. election cycle, people disseminated a fake story that the Pope endorsed Donald Trump, and almost a million people shared it.[33] Traditional media pointed out the falsity of the story, but many voters could not care less. Truth is no longer as important as seeming or feeling something to be true since "people often tune in to ideologically resonant sources of information,"[34] engage in confirmation bias, resist information that is inconsistent with their ideology, and promote their favorite narratives regardless of truth.

As technology advances, automation and artificial intelligence now allow for the creation of deepfakes—believable videos, photos, and audio of people doing and saying things they never did[35]—that

---

33. Zeynep Tufekci, Opinion, *Mark Zuckerberg Is in Denial*, N.Y. TIMES (Nov. 15, 2016), https://www.nytimes.com/2016/11/15/opinion [https://perma.cc/6S5X-8HXM]; ZEYNEP TUFECKI, TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST 264–65 (2017) [hereinafter TUFEKCI, TWITTER AND TEAR GAS]

34. TUFECKI, TWITTER AND TEAR GAS, *supra* note 33, at 40.

35. Robert Chesney & Danielle Citron, *Deepfakes: A Looming Crisis for National Security,* LAWFARE (Feb. 21, 2018), https://www.lawfareblog.com/deepfakes-looming-crisis-national-security-democracy-and-privacy# [https://perma.cc/KZ97-K9KP] ("Machine-learning algorithms . . . combined with facial-mapping software enable the cheap and easy fabrication of

generates even greater manipulation of the truth.[36] Liars can easily avoid accountability, claiming that true statements are fake stories. In contrast, truth-tellers can be portrayed as liars.[37] Falsehoods spread faster than truth, grab the attention of the audience, and enhance dissemination.[38] In this environment, false rumors have particular importance. The internet enables these rumors to spread even faster than previously and cause great harm to the reputations of individuals in particular and society in general. But what are the possible motives for publishing falsehoods in the first place?

1. Motivations for Publishing Falsehoods

Propagators of false rumors have diverse motivations.[39] They can spread falsehood intentionally, negligently, or recklessly.[40] There are four principle motives for publishing falsehoods.[41] Some propagators are *narrowly* self-interested: by spreading falsehoods, they

---

content that hijacks one's identity—voice, face, body."); *see also* Sunstein, *supra* note 8, at 387, 419.

36. *See* Robert Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753, 1759–60 (2019) (explaining the emergence of machine learning through neural network methods that increase the capacity to create false images, videos, and audio and that generative adversarial networks can lead to the production of increasingly convincing and nearly impossible to debunk deep fakes); Sunstein, *supra* note 8, at 423. Neural networks can also be used for AI creation of news stories that mimic the style and substance of real news stories. *See* Sarah Kreps, *The Role of Technology in Online Misinformation*, BROOKINGS: FOREIGN POLICY 1, 6 (2020), brook.gs/2Y6KzL1.

37. *See* Lili Levi, *Real "Fake News" and Fake "Fake News,"* 16 FIRST AMEND. L. REV. 232, 234 (2018) ("In that spirit, President Trump has deployed the 'fake news' trope to demonize and dismiss the traditional press as the 'enemy of the American people.' "); Chesney & Citron, *supra* note 36, at 1785 (describing that difficulty in separating truth from falsehood allows a "liar's dividend" because anyone can claim that a true story is fake while his lies are the truth).

38. Vosoughi et al., *supra* note 15, at 1147.

39. ANDREW MARANTZ, ANTISOCIAL: ONLINE EXTREMISTS, TECHNO-UTOPIANS, AND THE HIJACKING OF THE AMERICAN CONVERSATION 120 (2019) ("[Y]ou can post something because you believe it, or because you didn't believe it and you wanted to see who would. You can post something because you valued freedom of thought for its own sake; you can post something solely to get a reaction; you can post something without even knowing why, just because you felt like it.").

40. BENKLER ET AL., *supra* note 2, at 23–37 (differentiating between different types of dissemination of political fake news under the umbrella of propaganda and dividing it into four types: 1) manipulation- "directly influencing a person's beliefs, attitudes, or preferences in ways that fall short of what an empathetic observer would deem normatively appropriate in the context"; 2) misinformation- "communication of false information without intent to deceive, manipulate, or otherwise obtain an outcome"; 3) disinformation- "dissemination of explicitly false or misleading information"; and 4) bullshit- "commercial actors with no apparent political agenda who propagate[] made-up stories to garner [business] engagements and advertising revenue" and are indifferent to whether the stories are true or false).

41. *See* CASS R SUNSTEIN, ON RUMORS: HOW FALSEHOODS SPREAD, WHY WE BELIEVE THEM, WHAT CAN BE DONE 12–15 (2009).

aim to promote their own interest. In other words, they aim to harm a particular person, group, or competitor in the political or commercial realm and promote themselves by degrading their rivals.[42] "Other propagators are *generally* self-interested[: t]hey . . . seek to attract [an audience] by spreading rumors."[43] In contrast to the narrowly self-interested, generally self-interested propagators are indifferent to whether the rumor is true or false. Another type of propagator is actually *altruistic* and disseminates rumors they believe to be true without checking the facts.[44] Finally, *malicious propagators* intentionally seek to disseminate damaging information about individuals and institutions simply to inflict harm.[45]

2. Why Do People Disseminate Falsehoods?

An initial post may have a limited number of recipients; however, these recipients may then share it with others, leading to extensive dissemination and severe harm. Initiating a rumor is one thing, but what makes another person spread it? Or, why do people share information in general?

The nature of the internet's social environment fuels the distribution of ideas, information, and rumors at minimal cost. Constant connection to the internet allows anyone to share information. Thus, an idea can spread exponentially and reach a global audience at the click of a button.[46] Furthermore, as a falsehood circulates, it tends to become more menacing since the more individuals are exposed to a particular statement, the more likely they are to believe it and perceive it as a known fact.[47] Additionally, on the

---

42. For instance, individuals may spread rumors in order to make money, win a competition or political race, including by spreading negative, fake stories about competitors, or otherwise to get ahead. *See id.*; *see also, e.g.*, Ronan Bergman, *Twitter Network Uses Fake Accounts To Promote Netanyahu, Israel Watchdog Finds,* N.Y. TIMES (Mar. 31, 2019), https://www.nytimes.com/2019/03/31/world/middleeast/netanyahu-fake-twitter.html [https://perma.cc/XQ6M-36X8.

43. SUNSTEIN, *supra* note 41, at 13 (emphasis added); *see* JACOB SILVERMAN, TERMS OF SERVICE: SOCIAL MEDIA AND THE PRICE OF CONSTANT CONNECTION 45 (2015) (explaining that an important reason for publishing information is that people want to show others they are active on the social network and get their attention).

44. *See* SUNSTEIN, *supra* note 41, at 13.

45. *See id.* at 14.

46. The internet simplifies the dissemination of information and allows sharing with a large audience with the click of a button. *See generally* DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE (2014); LEE RAINIE & BARRY WELLMAN, NETWORKED: THE NEW SOCIAL OPERATING SYSTEM 67 (2012); DAVID A POTTS, CYBERLIBEL: INFORMATION WARFARE IN THE 21ST CENTURY 30 (2011); Jacqueline D. Lipton, *"We, the Paparazzi": Developing a Privacy Paradigm for Digital Video*, 95 IOWA L. REV. 919 (2010).

47. CASS R. SUNSTEIN, CONSPIRACY THEORIES AND OTHER DANGEROUS IDEAS 25–27 (2014); NICHOLAS DIFONZO & PRASHANT BORDIA, RUMOR PSYCHOLOGY: SOCIAL AND

internet, word-of-mouth information is not ephemeral. It remains accessible indefinitely through a Google search.[48] The more widely shared a falsehood is, the higher it appears in Google and other search engines' results, leading to even greater exposure and causing users to ascribe it more and more relevance.[49] Thus, the dissemination of a falsehood has the potential to cause tremendous harm to a person's reputation.[50]

Yet, not all falsehoods are spread as extensively as others; some are only disseminated locally. Why do some falsehoods spread widely while others remain limited in reach? In his seminal work *Threshold Models of Collective Behavior*, Mark Granovetter introduced the idea of "threshold" and maintains that it explains these processes of adoption of behavior.[51] Threshold refers to the proportion of the group that needs to join an activity before an individual follows suit.[52] Thus, "[o]ne's social network has a huge potential to affect one's decisions to adopt and disseminate certain ideas"[53] because people respond to the influences and preferences of others.[54]

A well-known U.S. election study serves as a good example.[55] Many Facebook users were shown a button to click to indicate that they had voted. Clicking the button created and shared a post about voter participation. For some users, the post included a graphic sign and pictures of friends in the social network who also indicated they had voted. "Researchers cross-referenced names with actual voting records and found that those people who saw

---

ORGANIZATIONAL APPROACHES 225 (2007); Pennycook et al., *supra* note 16 (explaining that the more people hear information, the more likely they are to believe it and pass it on).

48. *See* DANAH BOYD, IT'S COMPLICATED: THE SOCIAL LIVES OF NETWORKED TEENS 11 (2014); Daniel J. Solove, *Speech, Privacy, and Reputation on the Internet*, *in* THE OFFENSIVE INTERNET 15, 15–19 (Saul Levmore & Martha C. Nussbaum eds., 2010); Lavi, *supra* note 32, at 2603.

49. *See* SIVA VAIDHYANATHAN, THE GOOGLIZATION OF EVERYTHING (AND WHY WE SHOULD WORRY) 20–21 (2011); Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay & Laura Granka, *In Google We Trust: Users' Decisions on Rank, Position, and Relevance*, 12 J. COMPUT.-MEDIATED COMMC'N 801 (2007); Lavi, *supra* note 32, at 2604.

50. *See* CITRON, *supra* note 46, at 197–99.

51. Mark Granovetter, *Threshold Models of Collective Behavior*, 83 AM. J. SOCIO. 1420, 1422 (1978).

52. *Id.*

53. Michal Lavi, *Evil Nudges*, VAND. J. ENT. & TECH. L. 1, 16 (2018).

54. *See* NICHOLAS A. CHRISTAKIS & JAMES H. FOWLER, CONNECTED: THE SURPRISING POWER OF OUR SOCIAL NETWORKS AND HOW THEY SHAPE OUR LIVES 127 (2009); Michal Lavi, *Content Providers' Secondary Liability: A Social Network Perspective*, 26 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 855, 889 (2016).

55. Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle & James H. Fowler, *A 61-Million-Person Experiment in Social Influence and Political Mobilization*, 489 NATURE 7415 (2012), https://www.nature.com/articles/nature 11421. For a discussion of the study and its results, see Jonathan Zittrain, *Engineering an Election,* 127 HARV. L. REV. F. 335, 335–36 (2014).

posts [indicating] that their friends [had] voted were more likely to vote" themselves.[56]

In addition to collective thresholds, each individual has his own personal threshold for adopting and disseminating ideas.[57] In this context, one can identify three types of individuals. "Receptives" are individuals who are already disposed to favor a newly presented idea or share the same ideology.[58] Therefore, they have the lowest threshold and tend to adopt information they receive and pass it on.[59] Another group of disseminators is the "neutrals." This group has no inclination in favor or against an idea. If they notice that a few people have accepted and disseminated an idea, they may come to accept it and disseminate it as well.[60] Finally, there are the "skeptics," who have a high threshold for accepting and disseminating ideas and may have a prior disposition against certain ideas. Skeptics require a great deal of information before accepting new ideas. Once the evidence becomes overwhelming—and this evidence may include beliefs shared by many others—the skeptics will follow suit and accept the idea.[61]

Because individuals influence one another, ideas, including falsehoods, can spread through informational cascades.[62] In other words, people disseminate falsehoods, because others previously disseminated them, without holding a prior disposition or ideology that supports them. When an increasing number of people believe a falsehood, it can begin to appear credible and consequently influence others to believe it as well. Social pressure can also push people to spread information. In such cases, "people think they know what is right, or what is likely to be right, but they nonetheless go along with the crowd in order to maintain their status."[63]

---

56. Lavi, *supra* note 53, at 15 n.79; Lavi, *supra* note 25, at 147 n.8; *see also* Zittrain, *supra* note 55, at 335–36.

57. Granovetter, *supra* note 51, at 1423.

58. *See* SUNSTEIN, *supra* note 41, at 19–20; Edward Glaeser & Cass R. Sunstein, *Does More Speech Correct Falsehoods?*, 43 J. LEGAL STUD. 65, 67 (2014) (explaining that people have different prior beliefs and hence different degrees of skepticism and that "[i]ndividuals who believe that the messenger is a truth teller" tend to give credence to their statements).

59. *See* SUNSTEIN, *supra* note 41, at 19 (explaining that the individual threshold depends on a person's prior disposition regarding the information).

60. *See id.* at 20.

61. *See id.*; Lavi, *supra* note 53, at 17.

62. "Informational cascades are generated when individuals follow the statements or actions of predecessors and refrain from expressing opposing opinions because they believe their predecessors are right." Lavi, *supra* note 32, at 2602 n.14. As a result, important information is omitted from the social network. *See* Cass R. Sunstein & Reid Hastie, *Four Failures of Deliberating Groups* 2 (Univ. of Chi. Pub. L., Working Paper No. 215, 2008); *see also, e.g.*, Matthew Salganik, Peter Dodds & Duncan Watts, *Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market*, 311 SCI. MAG. 854 (2006).

63. Lavi, *supra* note 53, at 18 n.107; *see also* CASS R. SUNSTEIN, INFOTOPIA: HOW MANY MINDS PRODUCE KNOWLEDGE 91 (2006); Sunstein & Hastie, *supra* note 62.

This is the phenomenon of reputation cascades. As the crowd grows, the risk that a large number of people will believe entirely false information becomes reality.[64] Diffusion of ideas, trends, or behavior starts slowly. The idea can, however, spread rapidly "when a critical mass of individuals publicly share[s the] idea [and] a 'tipping point' occurs."[65]

The spreading and adoption of a rumor depends on encountering individuals with low thresholds who are willing to spread it further. When individuals are surrounded by peers with similar low thresholds, the diffusion of a rumor can accelerate. Yet, overlapping social circles might create structural holes and hinder dissemination.[66] It is difficult, however, to predict these tipping points when ideas are widely spread, as every individual in the network has a different threshold and they operate in different types of networks that can be random or homogenous.[67] The threshold is reliant on a variety of personal elements and social structures.[68] Changes in the social network's composition, social structures, and the transition path of an idea can significantly change the likelihood of widespread dissemination.[69] Social networks have a tremendous impact on the flow of information. They can withhold or accelerate the dissemination of rumors and this is the key to understanding how information and, in particular, falsehoods are disseminated.[70]

Having presented the threshold as an important factor influencing the spread of rumors, the next section addresses the central factors impacting the likelihood of reaching this threshold: the strength of social network ties and the influence of online intermediaries.[71]

---

64. *See* SUNSTEIN, *supra* note 41, at 2.

65. Lavi, *supra* note 53, at 17; *see also* MALCOLM GLADWELL, THE TIPPING POINT: HOW LITTLE THINGS CAN MAKE A BIG DIFFERENCE 12 (2002) (defining a tipping point as "the moment of critical mass, the threshold, the boiling point"). These principles were demonstrated in many contexts such as diffusion of innovation. *See* EVERETT M ROGERS: DIFFUSION OF INNOVATION (5th ed. 2003).

66. *Cf.* CHARLES KADUSHIN, UNDERSTANDING SOCIAL NETWORKS: THEORIES, CONCEPTS AND FINDINGS 158 (2012).

67. *See* Granovetter, *supra* note 51, at 1423 (demonstrating this point by using diffusion of rumors).

68. Lavi, *supra* note 53, at 17; *see also* KADUSHIN, *supra* note 66, at 156, 160–61 (2011).

69. *See* KADUSHIN, *supra* note 66, at 159–61.

70. *Cf.* CHRISTAKIS & FOWLER, *supra* note 54, at 3–32.

71. It should be noted that there are additional factors such as the interference of foreign countries, hackers, and the norms of traditional media in the context of fake news. Yet, the main factors are "bottom-up" network structures and "top-down" influences of intermediaries on networks including manipulation through algorithms. For further information, see BENKLER ET AL. *supra* note 2, at 8–23, which explains how network architecture, including a right-wing website that repeated fake news, allowed and exacerbated the spread of rumors ascribing corruption to Hillary Clinton during the 2016 U.S. election campaign and

B. *Connected: Network, Ties, and Social Structures*

A social network is a set of relationships.[72] These relationships structure the flow of interactions and social life today.[73] They are always present, influencing our choices, actions, thoughts, feelings, and desires.[74] Social networks shape norms and impact every aspect of the human experience.[75]

Network theory explains how connections between discrete objects are created, develop, and change.[76] Sociologists of networks focus on the influences of networks on communication patterns and the ties between individuals rather than on what a given individual thinks or does independently.[77] These ties shape interactions among members of social networks and, therefore, show how information is disseminated within social networks.[78] Thus, studying social networks can provide an enlightened understanding of social dynamics and information dissemination.[79] Understanding social networks makes it possible to explain the flow of information and is the first step towards outlining information policy for online dissemination of rumors.

Sociologists have addressed how different types of ties on social networks influence the flow of information.[80] When information is transmitted in a network characterized by strong ties (such as close

---

found that the structure of social networks was more important than intervention by Russian hackers, Facebook algorithms, online echo chambers, or Cambridge Analytica. However, these findings focused on the 2016 U.S. election campaign and do not purport to reach general conclusions regarding the most influential factors for spreading false rumors on the whole. Moreover, network structure and intermediaries both have an impact in directing the flow of information; the factors influence each other, and both have significant sway on dissemination.

72. *See* CHRISTAKIS & FOWLER, *supra* note 54, at 9.

73. Lavi, *supra* note 54, at 889; *see also* Manuel Castells, *Afterword: Why Networks Matter*, *in* NETWORK LOGIC 219, 221 (2004). On networks in general, see also JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 40 (2019) ("A *network* is a mode of organization in which hubs and nodes structure the flows of transactions and interactions.").

74. Lavi, *supra* note 54, at 889.

75. *Id.*

76. *Id.* at 890; *cf.* KADUSHIN, *supra* note 66, at 27. Kadushin's book focuses on social networks and sociological theory. For a wider understanding of network science, see generally DUNCAN J. WATTS, SIX DEGREES: THE SCIENCE OF A CONNECTED AGE (2004); and ALBERT-LÁSZLÓ BARABÁSI, LINKED: THE NEW SCIENCE OF NETWORKS (2014).

77. Lavi, *supra* note 54, at 890; *see also* BARABÁSI, *supra* note 76; KADUSHIN, *supra* note 66, at 27; RAINIE & WELLMAN, *supra* note 46, at 42; *cf.* Caroline Haythornthwaite, *Social Networks and Internet Connectivity Effects*, 8 INFO. COMMC'N. & SOC'Y, 125, 127 (2005).

78. Lavi, *supra* note 54, at 889–90; *see also* CHRISTAKIS & FOWLER, *supra* note 54, at 7–9.

79. Lavi, *supra* note 54, at 890.

80. *See e.g.*, Mark Granovetter, *The Strength of Weak Ties*, 78 AM. J. SOCIO. 1360, 1361 (1973); Mark Granovetter, *The Strength of Weak Ties: A Network Theory Revisited*, 1 SOCIO. THEORY 201 (1983) [hereinafter Granovetter, *A Network Theory Revisited*]; Ronald S. Burt, *The Network Structure of Social Capital*, 22 RSCH. ORGANIZATIONAL BEHAV. 345, 345–49, 353, 359 (2000).

friends and family), the recipient ascribes credibility to the information because he knows and trusts its source.[81] In addition, interdependency between people increases social pressure and the likelihood of crossing the threshold and disseminating an idea.[82] Information transmitted in this kind of network tends to spread quickly among followers but slows down outside the cluster of ties due to overlapping social circles that hinder dissemination between distant parts of the social network.[83] In contrast, weak ties facilitate fast dissemination of information. In fact, they can bridge the structural holes between non-overlapping clusters of strong ties.[84] Yet, individuals who receive information via weak ties may ascribe it less weight. [85]

The following two subsections focus on the dissemination of rumors on social network platforms. Section I.C describes briefly the social structures of these platforms and their "bottom-up" influences on dissemination, and Section I.D describes the "top-down" influences of intermediaries on these platforms. Due to these dynamics, it argues that social network platforms lend a high probability of users meeting their individual thresholds for disseminating falsehoods.

C. *Social Network Platforms and Dissemination of Rumors: Strength of Ties, Thresholds, and "Bottom-up" Influence on Social Dynamics*

More than twenty years ago, sociologists Gustavo Mesch and Ilan Talmud mapped three social factors affecting the quality of online ties: (1) social similarity (homophily); (2) the intensity of contact (relationship duration); and (3) different dimensions of the relationship (multiplexity).[86] An application of these factors to social

---

81. David Krackhardt, *The Strength of Strong Ties: The Importance of Philos in Organizations*, *in* NETWORKS AND ORGANIZATIONS: STRUCTURE, FORM AND ACTION 216, 218 (N. Nohria & Robert G. Eccles eds., 1992) ("Strong ties constitute a base of trust that can reduce resistance and provide comfort in the face of uncertainty.").

82. *See* Lavi, *supra* note 53, at 17; *cf.* Ronald S. Burt, *Social Contagion and Innovation: Cohesion Versus Structural Equivalence*, 92 AM. J. SOC. 1287, 1290 (1987); Mark Granovetter & Roland Soong, *Threshold Models of Diffusion and Collective Behavior*, 9 J. MATH. SOC. 165, 165–66 (1983) (focusing on the homogeneity assumption in models where the network is composed of homogenous individuals).

83. *See* Lior Jacob Strahilevitz, *A Social Networks Theory of Privacy*, 72 U. CHI. L. REV. 919, 956 (2005) ("Information transmitted via strong ties generally spreads less quickly, but is more accurate and credible.").

84. TUFECKI, TWITTER AND TEAR GAS, *supra* note 33, at 21–22 (explaining how a mixture of strong and weak ties impacts the diffusion of protest movements)

85. Granovetter, *A Network Theory Revisited, supra* note 80, at 218–19; Strahilevitz, *supra* note 83, at 965; Krackhardt, *supra* note 81, at 218; KADUSHIN, *supra* note 66, at 69;.

86. Gustavo S. Mesch & Ilan Talmud, *The Quality of Online and Offline Relationships: The Roles of Multiplexity and Duration of Social Relationships*, 22 INFO. SOC'Y 137 (2006).

network platforms reveals that these platforms facilitate the formation of strong ties.

Social network platforms, such as Facebook and Twitter, are "web-based services that allow individuals to: (1) construct a public or semi-public profile within a bounded system; (2) articulate a list of other users with whom they share a connection; and (3) view and traverse their list of connections and those made by others within the system."[87] The design and interfaces of the network have significant implications on the types of interactions.[88] The design allows people to cluster around similar users who share a common denominator and, thus, participants are relatively homogenous within clusters. Moreover, social network platforms allow continuing interactions among repeat players, each of whom has a personal profile that represents their real identity in the physical world.[89] Finally, people discuss diversified and personal subjects on social network platforms. Thus, strong ties are likely to form among participants of social network platforms, in contrast to other types of platforms,[90] and these ties have an extensive "bottom-up" effect on the flow of information.

Strong social ties influence a recipient's perception of speech. Information transmitted via strong ties may be complex, personal, and perceived as more credible.[91] Due to the likelihood of strong ties on social network platforms, the harm to reputation and public interest in this setting can be extensive. First, in most social networks, speech is not anonymous. The personal profile a user creates within the system usually represents his real identity. The source of the message is known and, thus, the message is perceived as more credible than if it were to come from an anonymous source.[92] Second, social similarity and homogeneity on social net-

---

87. James Grimmelmann, *Saving Facebook*, 94 IOWA L. REV. 1137, 1142 (2009) (citing Danah Boyd & Nicole Ellison, *Social Network Sites: Definition, History, and Scholarship*, 13 J. COMPUT.-MEDIATED COMMC'N 210, 211 (2008)).

88. *See* IAN BROWN & CHRISTOPHER T. MARSDEN, REGULATING CODE: GOOD GOVERNANCE AND BETTER REGULATION IN THE INFORMATION AGE 118–19 (2013); Grimmelmann, *supra* note 87, at 1143.

89. *Cf.* Grimmelmann, *supra* note 87, at 1198 (explaining that Facebook has discretion to remove users who do not use their real identity).

90. On social networks, strong ties are likely to form due to overlapping social circles and similarity among users who cluster around one another, continuous two-way interaction between repeat players, and various complex subjects of discussion. *See* Lavi, *supra* note 54, at 903. Strong ties are less likely to form on other types of platforms. *Id.* at 893–94. Drawing on network theory, especially the key factors affecting the quality of social ties, I outlined a descriptive taxonomy of online platforms. I identified three categories of platforms based on the strength of ties formed between their users: (1) freestyle discourse; (2) peer-production and (3) deliberation and structuring communities (including social networks). *Id.* at 895–908.

91. *See* Lavi, *supra* note 54, at 890–91.

92. *Id.* at 893–94; Grimmelmann, *supra* note 87, at 1198.

works may lead to interdependence among participants,[93] and their social network influences their decisions.[94] Clusters of strong ties can create echo chambers where similar individuals support and echo information, confirming prior convictions and entrenching their credibility, while voiding other viewpoints.[95] Such "bottom-up" influence of strong ties on social dynamics increases the likelihood that members will reach their thresholds for accepting and spreading information.[96] Third, users may add comments to the shared content, thereby validating and reinforcing it. Thus, the process of social sharing has "bottom-up" influences that amplify the gravity of harm.[97]

On social network platforms, content can spread by word-of-mouth among actors who know each other, interact, or have mutual interests.[98] It can also spread by imitation via "memes"[99] or information cascades.[100] When many people simultaneously forward a specific item of information over a short period of time and it spreads beyond their own social network, the content becomes viral.[101]

Further, a message shared through social networks may be more influential than a message shared by mass media.[102] For example, "friends" on social networks generate similar types of content.[103]

---

93. *See* Granovetter & Soong, *supra* note 82 (referring to the homogeneity assumption).

94. James H. Fowler & Nicholas A. Christakis, *Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study*, 337 BMJ a2338 (2008); Nicholas A. Christakis & James H. Fowler, *The Spread of Obesity in a Large Social Network over 32 Years*, 357 NEW ENG. J. MED. 370–79 (2007).

95. CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA, 11, 116–135, 155 (2017); *see also* BENKLER ET AL. *supra* note 2, at 80–83 (expanding on the confirmation bias in the right wing media outlet that affected the dissemination of fake news in the 2016 U.S. election campaign).

96. *See* SUNSTEIN, *supra* note 41, at 20 (noting that with the shared view of a few people, they might come to accept the rumor).

97. *Id.* at 40; CASS R. SUNSTEIN & REID HASTIE, WISER: CONSPIRACY THEORIES 85 (2015).

98. *See* KARINE NAHON & JEFF HEMSLEY, GOING VIRAL 36 (2013).

99. *See* AN XIAO MINA, MEMES TO MOVEMENTS HOW THE WORLD'S MOST VIRAL MEDIA IS CHANGING SOCIAL PROTEST AND POWER 6, 20 (2019); LIMOR SHIFMAN, MEMES IN DIGITAL CULTURE 2, 9–15 (2013) (explaining how the term "meme" was coined by Richard Dawkins in 1976 to describe small units of culture that spread from person to person through copying or imitation; internet memes are posts in which shared norms and values are constructed through cultural artifacts) (referring to RICHARD DAWKINS, THE SELFISH GENE (1976)).

100. NAHON & HEMSLEY, *supra* note 98, at 38–39; *see also* Burt, *supra* note 82, at 1290 (referring to imitation and stickiness that increase the likelihood of diffusing information).

101. NAHON & HEMSLEY, *supra* note 98, at 16.

102. *See* ELIHU KATZ & PAUL F. LAZARSFELD, PERSONAL INFLUENCE: THE PART PLAYED BY PEOPLE IN THE FLOW OF MASS COMMUNICATION 32 (1955) (explaining that most people express their opinions under the influence of central hubs in their social networks (the "opinion leaders")).

103. In a scientific experiment, "Facebook showed some users fewer of their friends' posts containing emotional language [and] then analyzed the users' own posts to see whether their emotional language changed." James Grimmelmann, *The Law and Ethics of Experiments on Social Media Users*, 13 COLO. TECH. L.J. 219, 222 (2015) (citing Adam D.I.

The strength of ties and context formed on social network platforms allows "top-down" social dynamics that enhance social pressures.[104] It also increases the likelihood that multiple users will reach their threshold for accepting and disseminating speech through their social ties. Thus, even neutral individuals who do not have a prior disposition to support information may adopt and spread a rumor. Moreover, even skeptics, who see that other individuals in their social network have disseminated a rumor, may eventually reach their threshold for adopting and disseminating it.[105] Due to the strong ties on social networks, the likelihood for correcting false rumors, defamation, and fake news decreases.[106]

Indeed, not all ties in a social setting are strong, and alongside strong ties, there are weak ties. These weak ties create a bridge between structural holes in the network and allow for vast dissemination. As a result, in online social networks, information is perceived as credible because of the strong ties and may travel between distant parts of the social network, disseminating vastly over a short period of time due to weak ties that bridge between non-overlapping clusters of strong ties.[107] Thus, weak ties accelerate the dissemination of information beyond clusters of strong ties, whereas strong ties enhance the credibility of information.

D. *The "Top-down" Influence of Online Intermediaries on Networks and Thresholds*

The social dynamics of content dissemination are not the whole story. In addition to "bottom-up" dynamics, intermediaries of social network platforms influence users from the "top-down," causing users to cross their thresholds and disseminate ideas.[108]

Kramer, Jamie E. Guillory & Jeffrey T. Hancock, *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks*, 111 PROC. NAT'L ACAD. SCIS. U.S.A. 8788, 8788 (2014)). A change was indeed found. *Id.*

104. *See* CASS R. SUNSTEIN & REID HASTIE, WISER: GETTING BEYOND GROUPTHINK TO MAKE GROUPS SMARTER 85–86 (2014).

105. NAHON & HEMSLEY, *supra* note 98, at 33 (explaining that the effects of the social network enhance the dissemination of information).

106. *See* Lavi, *supra* note 54, at 918; SUNSTEIN & HASTIE, *supra* note 97, at 25–27 (explaining that in some cases efforts to correct a false rumor or conspiracy theory can have the opposite effect and even cause more people to believe it).

107. *See* KADUSHIN, *supra* note 66, at 69 (explaining that since weak ties form among acquaintances, they facilitate the dissemination of information beyond the cluster of strong ties).

108. NEIL RICHARDS, INTELLECTUAL PRIVACY: RETHINKING DIGITAL LIBERTIES IN THE DIGITAL AGE 170–74 (2015); Danielle Keats Citron & Helen Norton, *Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age*, 91 B.U. L. REV. 1435, 1456 (2011); *see also* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 60 (2015); James Grimmelmann, *The Virtues of Moderation*, 17 YALE

Individuals operate in a social context, and their behavior depends on the architecture of their environment. The technological design of platforms has political dimensions.[109] Values can be "baked into" the technological architecture, offering a "seductively elegant and effective means of control."[110] Each and every choice of architecture affects the social network's context and interpersonal dynamics between users.[111] Intermediaries can influence decisions to generate and disseminate content through their choice of architecture.[112] Such a choice also creates "affordances"—the possible actions and uses that can be performed on the platform. These affordances, in turn, influence how people use the platform.[113] Just a few tweaks in the design of an intermediary's platform can make a huge difference in how it is used and, consequently, its potential for the widespread circulation of ideas. Understanding social dynamics on social networks allows intermediaries to harness technology and design their platform to influence the flow of information from the top down.

Intermediaries earn more revenue from advertisers as participation increases, since social engagement keeps users on the platform longer. Continued participation allows intermediaries to collect more information on users,[114] monetize "social graph," target personalized advertisements, and maximize profits.[115] Therefore,

---

J.L. & TECH. 42, 55 (2015) (expanding on governing mechanisms that structure participation in a community to facilitate cooperation).

109. *See* Langdon Winner, *Do Artifacts Have Politics?*, 109 DAEDALUS 121 (1980) (discussing in detail the political dimensions of technological design).

110. Ari Ezra Waldman, *Privacy's Law of Design*, 9 U.C. IRVINE L. REV. 1239, 1242 (2019) ("[D]esign's awesome yet invisible capacity to manipulate those who exist inside its ecosystem requires us to consider the values we want design to promote."); Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-By-Design*, 106 CAL. L. REV. 697, 701, 721 (2018); *see also* Woodrow Hartzog, *Body Cameras and the Path to Redeem Privacy Law*, 96 N.C. L. REV. 1257, 1299 (2018) (discussing the value in body cameras that track police officers' behavior).

111. *See* NAHON & HEMSLEY, *supra* note 98, at 82 (explaining that Twitter limits the number of characters in tweets. The platform was intentionally structured this way by its designers. Consequently, people use it for short reports on what they are doing); TUFECKI, TWITTER AND TEARGAS, *supra* note 33, at 267 (noting the possibilities for technology to shape and influence the dissemination of rumors).

112. *See* Lavi, *supra* note 53, at 14. Technology plays an important role in influencing contexts. *See, e.g.*, NAHON & HEMSLEY, *supra* note 98, at 82. (giving an example of how character limitations influence the flow of information). *See generally* B.J. FOGG, PERSUASIVE TECHNOLOGY: USING COMPUTERS TO CHANGE WHAT WE THINK AND DO 5 (2003).

113. RAINIE & WELLMAN, *supra* note 46, at 65; Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904, 1913 (2013) ("[O]ur artifacts organize the world for us, subtly shaping the ways that we make sense of it.")

114. COHEN, *supra* note 73, at 65, 83 ("Platform-based, massively intermediated environments enable people seeking connection with each other to signal their affinities and inclinations using forms of shorthand—'Like', 'Follow,' 'Retweet,' and so on—that simultaneously enable data capture and extraction.").

115. *Id.* at 55; MARY ANNE FRANKS, THE CULT OF THE CONSTITUTION 171 (2019) ("The more content users voluntarily provide (posts, shares, likes etc.), the more users interact on

intermediaries strive to maximize the participation of users and social sharing.[116] To accomplish this, they use insights gleaned from sociology, psychology, and management.[117] "These insights allow intermediaries to predict cognitive biases and social dynamics, deploy new socio-technical systems, and influence flows of information."[118] Such influences are operating from the top down.[119] Similar to the gaming industry, design and technology can turn the use of social media addictive.[120]

For example, social network intermediaries bolster user motivation to spread content, make it easier for them to share information, and trigger them to do so.[121] They excel in making users feel socially connected. Pictures, names, and other informal signs create the feeling that mere contacts are close friends.[122] Moreover, intermediaries frequently "utilize algorithms to prioritize newsfeed content created by a user's close friends and family, which rein-

---

the platform, and the more companies like Facebook can target users with increasingly personal advertising. If harmful content provided by a user generates a high level of engagement from a large number of users, then the advertising benefit of that post goes up, which means more money in Facebook's pocket."); SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019) (coining the term "surveillance capitalism" to describe tracking users' engagements in order to enhance commercial profits); 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, 1555 (2018); Danielle Keats Citron, *Cyber Mobs, Disinformation, and Death Videos: The Internet as It Is (and as It Should Be)*, 118 MICH. L. REV. 1073, 1085–86 (2020) (reviewing NICK DRNASO, SABRINA (2018)).

116. *See* MARANTZ, *supra* note 39, at 80 ("Facebook's larger goal, which always went unstated, was not to spread high-quality content; it was to entice more users into spending more time on Facebook."); JOSEPH TUROW, THE DAILY YOU: HOW THE NEW ADVERTISING INDUSTRY IS DEFINING YOUR IDENTITY AND YOUR WORTH 72, 102 (2012); ARI EZRA WALDMAN, PRIVACY AS TRUST: INFORMATION PRIVACY FOR AN INFORMATION AGE 90 (2018); Julie E Cohen, *Law for the Platform Economy*, 51 U.C. DAVIS L. REV. 133, 140 (2017); Daniel J. Solove, *The Myth of the Privacy Paradox* 14 (George Wash. Legal Studies Research Paper No. 2020-10, 2020). *See generally* NICHOLAS CARR, THE BIG SWITCH: REWIRING THE WORLD FROM EDISON TO GOOGLE 154–57 (2009) (discussing the development of intermediary dynamics across various industries).

117. Lavi, *supra* note 53, at 12.

118. *Id.*

119. *See id.* at 12, 18–19.

120. ZUBOFF, *supra* note 115, at 466 ("[J]ust as ordinary consumers can become compulsive gamblers at the hands of gaming industry, behavioral technology draws ordinary young people into an unprecedented vortex of social information . . . ."); *see also* Karen Mettler, *A Lawmaker Wants to End 'Social Media Addiction' by Killing Features that Enable Mindless Scrolling*, WASH. POST (July 30, 2019), wapo.st/2KBQ3X5; WOODROW HARTZOG: PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES 198 (2018) (expanding on architecture that makes the platform sticky and causes users to become addicted to the engagement);

121. *See* HARTZOG, *supra* note 120, at 197; JARON LANIER, TEN ARGUMENTS FOR DELETING YOUR SOCIAL MEDIA ACCOUNTS RIGHT NOW 18 (2018). *See generally* FOGG, *supra* note 112, at 198 (referring to socio-technical tools for enhancing users' motivation and ability to spread content, which intermediaries use to trigger users to spread information).

122. Grimmelmann, *supra* note 87, at 1162–63.

forces existing biases and further encourages dissemination."[123] Consequently, users share content they would not have necessarily shared offline.[124] This choice in architecture frames relationships and increases the likelihood that an individual will reach his threshold for accepting and spreading content.[125] It should also be noted that intermediaries can influence the content of information that users share, as the Cambridge Analytica scandal has demonstrated.[126] This Part, however, will focus on the influence intermediaries have on user decisions to share more information, regardless of its content or topic.

Intermediaries on social network platforms allow users to create personal profiles and declare that other users are their "friends." This framing of relationships enhances social trust and motivates users to divulge and share personal information. Defining every connection as a "friend" increases the likelihood of reaching the threshold to adopt and disseminate information, even though not all connections are actually their friends in the traditional sense of the word. [127]

Another example of how architecture choices can impact dissemination is social mirroring. This strategy reflects a user's behavior back to them via their newsfeed, leading to implied feedback and enforcing group identity.[128] Social mirroring enhances the

---

123. Lavi, *supra* note 53, at 30 n.213 (citing SUNSTEIN, *supra* note 95, at 16). In a related context Sacha Baron-Cohen demonstrated how the algorithmic code and algorithmic recommendations it targets encourages users to share specific types of content. *See* Sacha Baron Cohen, *Read Sacha Baron Cohen's Scathing Attack on Facebook in Full: 'Greatest Propaganda Machine in History,'* THE GUARDIAN (Nov. 22, 2019), https://www.theguardian.com/technology/2019/nov/22/sacha-baron-cohen-facebook-propaganda [https://perma.cc/7YPM-R8C7].

124. *See* Lavi, *supra* note 53, at 6 n.14; Samantha L. Miller, *The Facebook Frontier: Responding to the Changing Face of Privacy on the Internet*, 97 KY. L.J. 541, 546 (2008); *cf.* NICHOLAS CARR, THE GLASS CAGE: AUTOMATION AND US 179–82 (2014) (explaining the psychological impacts of techniques used by technology companies to facilitate bonds with their users).

125. *See* James Grimmelmann, *Accidental Privacy Spills*, 12 J. INTERNET L. 3, 6 (2008); Daniel Solove, *Introduction: Privacy, Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1886–88 (2013).

126. *See* BENKLER ET AL., *supra* note 2, at 11; Sam Meredith, *Here's Everything You Need to Know About the Cambridge Analytica Scandal*, CNBC (Mar. 23, 2018, 9:21 AM), https://www.cnbc.com/2018/03/21/facebook-cambridge-analytica-scandal-everything-you-need-to-know.html.

127. Ari Ezra Waldman, *Privacy, Sharing, and Trust: The Facebook Study*, 67 CASE W. RES. L. REV. 19, 221–22 (2016); *see also* BERNARD E HARCOURT, EXPOSED: DESIRE AND DISOBEDIENCE IN THE DIGITAL AGE 86, 99–100 (2015); Ari Ezra Waldman, *Safe Social Spaces,* WASH. U. L. REV. 1535, 1565–66 (2019); Ari Ezra Waldman, *Cognitive Biases, Dark Patterns, and the "Privacy Paradox,"* CURRENT ISSUES PSYCH. (forthcoming, 2020).

128. *See* Meng Ma & Ritu Agarwal, *Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities*, 18 INFO. SYS. RES. 42, 58 (2007); *cf.* Joan Morris DiMocco, Anna Pandolfo & Walter Bender, *Influencing Group Participation with a Shared Display*, 6 PROC. A.C.M. CONF. ON COMPUT. SUPPORTED COOP. WORK 614, 619 (2004) (discussing the results of experiments into the impacts of group dynamics on social processes).

network's influence on the individual user and increases the user's likelihood of reaching the threshold to join his "friends" and disseminate a rumor.[129] The power of social mirroring has been proven in an experiment conducted by Facebook. The social network only displayed the negative posts of "friends," omitting positive ones. As a result, users created and shared negative posts at higher rates than other types of content.[130]

Intermediaries selectively influence the content users see in their newsfeed and content is not always presented chronologically. Intermediaries can use Artificial Intelligence algorithms to tailor a users' newsfeed to present relevant content and prioritize the content of close friends and family.[131] This strategy increases the visibility of such content and the motivation to spread the information, because it increases confirmation bias among homogenous participants and enforces their beliefs.[132]

Intermediaries also allow explicit feedback by facilitating mechanisms for voting and the formation of reputations.[133] These mechanisms bolster mutual influence within the social network and pave the way for extensive dissemination of ideas within the network.[134]

---

129. *See* ZUBOFF, supra note 115, at 21, 306 (addressing social pressures and architecture that replace politics and democracy (confluence) and describing how social networks create context and influence the engagement of individuals); Alessandro Acquisti, Leslie K. John & George Loewenstein, *The Impact of Relative Standards on the Propensity to Disclose,* 49 J. MKTG. RSCH. 160, 162 (2012).

130. BRETT FRISCHMANN & EVAN SELINGER, RE-ENGINEERING HUMANITY 117–18 (2008) (describing Facebook's cognition experiment, testing users' emotions).

131. SUNSTEIN, *supra* note 95, at 14 (" '[T]he goal of [the] News Feed is to show people the stories that are most relevant to them.' With that point in mind, why does Facebook rank stories in its News Feed? 'So that people can see what they care about first, and don't miss important stuff from their friends.' "); *see also* SIVA VAIDHYANATHAN, ANTI-SOCIAL MEDIA: HOW FACEBOOK DISCONNECTS US AND UNDERMINES DEMOCRACY 77–105 (2018).

132. SUNSTEIN, *supra* note 95, at 122–24; BENKLER ET AL., *supra* note 2, at 76 ("[Individuals] look for media outlets and politicians that will inform them as best as possible without suffering too much cognitive discomfort."); *see also* Julie E. Cohen, *Internet Utopianism and the Practical Inevitability of Law*, 18 DUKE L. & TECH. REV. 85, 88 (2019) ("Algorithmic processes optimized to boost click-through rates and prompt social sharing heighten the volatility of online interactions, and surveillant assemblages designed to enhance capabilities for content targeting and behavioral marketing create powerful — and easily weaponized — stimulus-response feedback loops."). *See generally* ELI PARISER, THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU 35–48 (2011) (providing one of the first warnings that algorithms show links that users are more likely to click on).

133. For example, the "like" button on Facebook.

134. *See* COHEN, *supra* note 73, at 85 ("[S]ocial networking as Facebook and microblogging platforms, such as Twitter function as de facto aggregators for a wide range of content and deliver feeds optimized to everything that is known or inferred about particular users' opinions and beliefs. By design, all of those algorithms incorporate feedback effects, and so their operation both reflects and continually reinforces the powerful economic motivation to pursue viral spread."); BETH SIMONE NOVECK, WIKI GOVERNMENT 80–82 (2009); FOGG, *supra* note 112, at 44; Chrysanthos Dellarocas, *The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms*, 49 MGMT. & SCI. 1407, 1418 (2003) (discussing studies on the impacts of online feedback loops on establishing linkages between disconnected networks).

Intermediaries can also function as social actors and reward users for certain kinds of activity.[135] For example, Twitter used to reply automatically to every user who re-tweets and disseminates content with the message: "very nice." The ability of intermediaries to function as social actors has expanded with the development of Artificial Intelligence and Machine Learning algorithms that allow the operation of social bots. Such algorithmic software programs that operate according to intermerdiaries' instructions can interact socially with users, enhance their trust in the communication of the intermediary, and increase the likelihood of disseminating ideas.[136]

Intermediaries not only enhance the motivation to disseminate content but also simplify the ability to do so. For example, "forward," "share," or "re-tweet" buttons facilitate quick dissemination at the click of a button. There is no need for users to undergo the cumbersome copy and paste process in order to spread content. Due to the low cost of sharing and disseminating information, it is more likely that individuals will cross their "threshold" and join individuals already engaged in dissemination of information.[137] Simplifying dissemination encourages users to share information intuitively and almost automatically,[138] bypassing reflective thinking about the consequences of dissemination.[139] This choice of architecture engineers social behavior and influences decision-making, by promoting the fast dissemination of information to a wide audience.[140]

---

135. *See* B.J. Fogg & Clifford Nass, *Silicon Sycophants: The Effects of Computers That Flatter*, 46 INT'L J. HUM. COMPUT. STUDS. 551, 559–60 (1997) (discussing intermediaries as social actors and persuasive socio-technical tools).

136. *See* Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer & Alessandro Flammini, *The Rise of Social Bots*, 59 COMMC'NS A.C.M. 96, 96 (2016) ("A social bot is a computer algorithm that automatically produces content and interacts with humans on social media, trying to emulate and possibly alter their behavior."); WALDMAN, *supra* note 116, at 141 (expanding on the social communication of bots that motivate people to waive privacy protections, as a result of technological design); JARON LANIER, TEN ARGUMENTS FOR DELETING YOUR SOCIAL MEDIA ACCOUNTS RIGHT NOW 55–58 (2018) ("If your extended peer group contains a lot of fake people, calculated to manipulate you, you are likely be influenced without even realizing it.").

137. SUNSTEIN, *supra* note 95, at 108.

138. Julie E. Cohen, *Tailoring Election Regulation: The Platform Is the Frame,* 4 GEO. L. TECH. REV. (forthcoming 2020) (manuscript at 8) (explaining that platform-based, massively intermediated information environments are not designed for eliciting automatic, precognitive interactions with online content and discussing the aspects of platform interfaces that are designed for automatic, habitual engagement).

139. *See* DANIEL KAHNEMAN, THINKING, FAST AND SLOW 237 (2011) (explaining two systems of thinking: intuitive thinking, "system 1," and deliberative analytic thinking, "system 2").

140. FRISCHMANN & SELINGER, *supra* note 130, at 235 ("The smart social media environment that has emerged in the past decade of which Facebook is an important part – encourages people to accept what is presented to them without pushing for reflection or deliberation.").

This Section focused on the intermediaries' top-down influence on engagement and dissemination of any type of content. Beyond the focus of this Article,[141] it should be noted that intermediaries can prioritize false information on users' newsfeeds because such content inspires surprise and enhances engagement.[142] In short, social dynamics influence content dissemination from the bottom up and social network platform intermediaries enhance dissemination from the top down, both pushing users past their thresholds for dissemination of information. Although it is difficult to predict exactly when ideas will spread widely, both the strength of ties on social networks and the intermediary's choice architecture make it more likely that rumors will spread on online social network platforms than on other types of platforms.

E. *Publish, Share, Re-tweet, and Repeat: Benefits and Challenges*

Sharing content online can have a snowball effect and compound dissemination. As content gains more attention, users ascribe more weight to it.[143] Information disseminated can include harmful content, false rumors, defamation, and fake news, all inflicting tremendous harm. Despite this, the dissemination of information online can afford many benefits.

First, the sharing of information online promotes freedom of speech and the rationales at the base of this constitutional right.[144] It promotes individual autonomy and approval for the speaker's

---

141. For a discussion about the dangerous results of algorithmic recommendations involving harmful content, see Michal Lavi, *Do Platforms Kill?*, 43 HARV. J.L. & PUB. POL'Y 477, 500–05 (2020). In a related context, algorithmic impact assessment was proposed to accommodate the problem of discrimination and other harmful effects of algorithmic biases. For an expansion on the topic, see Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong. (2019). For further discussion and criticism on the proposed bill, see Margot E. Kaminski & Andrew D. Selbst, Opinion, *The Legislation That Targets the Racist Impacts of Tech*, N.Y. TIMES (May 7, 2019), https://www.nytimes.com/2019/05/07/opinion/tech-racism-algorithms.html [https://perma.cc/F4FL-7BMA].

142. VAIDHYANATHAN, *supra* note 131, at 5–6; *see also* Pauline T. Kim, *Manipulating Opportunity*, 106 VA. L. REV. 867, 894 (2020); Mark Bergen, *YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant*, BLOOMBERG (Apr. 2, 2019), https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant [https://perma.cc/LU9A-SW5N]; Cohen, *supra* note 132 (describing intermediaries as "the greatest propaganda machine in history").

143. The more times individuals are exposed to information, the more they tend to believe it. *See* SUNSTEIN, *supra* note 47, at 21 ("[R]umors frequently spread through information cascades. The basic dynamic behind such cascades is simple: once a certain number of people appear to believe a rumor, others will believe it to, unless they have good reason to believe it is false."); DIFONZO & BORDIA, *supra* note 47, at 225, Lavi, *supra* note 53, at 20.

144. Jack Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 3–4 (2004); Lavi, *supra* note 25, at 179; Lavi, *supra* note 54, at 879.

way of life.[145] It also promotes a vibrant *marketplace of ideas* as it enhances content accessibility and the right to receive information.[146] Dissemination also promotes *democracy* because it helps to keep citizens informed about acts of government and guarantees that policy is reached intelligently,[147] as almost every member of Congress operates a social media account.[148] Copying and disseminating content "promotes democracy by literally putting information in citizens' hands."[149] It also enhances civic involvement and collective action to promote important social and political goals,[150] even enabling protest.[151] In addition, dissemination protects a participatory democratic culture by enhancing dialogue on information from a broader variety of sources[152] and allowing all members a fair chance to develop and share ideas within the communities to which they belong.[153]

Second, dissemination of content encourages novel ways of consuming, transacting, and making a living.[154] It promotes efficiency by reducing the cost of information searches and increasing the

---

145. *See* Joseph Raz, *Free Expression and Personal Identification,* 11 OXFORD J. LEGAL STUD. 303, 313–16 (1991) (focusing on the importance of free speech in promoting individual autonomy).

146. *See* JOHN STUART MILL, ON LIBERTY 5–9 (1869) (the search for truth ensures that every expression enters the marketplace of ideas); JOHN MILTON, AREOPAGITICA: A SPEECH FOR THE LIBERTY OF UNLICENSED PRINTING (1958). The search for truth theory was popularized by Oliver Wendell Holmes's famous dissenting opinion in *Abrams v. United States*, 250 U.S. 616, 630 (1919), in which he stated that "the best test for truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out." *See* Sanchez, *supra* note 24, at 314–16.

147. *See* ALEXANDER MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT 26 (1948) (discussing the rationale for promoting democracy).

148. *See* Packingham v. North Carolina, 137 S. Ct. 1730, 1725–36 (2017) ("On Facebook, for example, users can debate religion and politics with their friends and neighbors or share vacation photos. On LinkedIn, users can look for work, advertise for employees, or review tips on entrepreneurship. And on Twitter, users can petition their elected representatives and otherwise engage with them in a direct manner. Indeed, Governors in all 50 States and almost every Member of Congress have set up accounts for this purpose.").

149. Rebecca Tushnet, *Copy This Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It,* 114 YALE L.J. 535, 565 (2004) (on the importance of copying and republishing expressions in maintaining democracy in the related context of copyright).

150. For instance, the dissemination of content allows unconnected individuals to organize and promote important goals such as attending the funeral of a lone soldier or efficiently organizing a civil protest. *See* CLAY SHIRKY, HERE COMES EVERYBODY: THE POWER OF ORGANIZING WITHOUT ORGANIZATION 146–64 (2008); CLAY SHIRKY, COGNITIVE SURPLUS: CREATIVITY AND GENEROSITY IN A CONNECTED AGE 175 (2010).

151. TUFECKI, TWITTER AND TEAR GAS, *supra* note 33, at 264.

152. Sanchez, *supra* note 24, at 316–17; *see also* Michael D. Birnhack, *More or Better? Shaping the Public Domain*, *in* THE FUTURE OF THE PUBLIC DOMAIN 59, 71 (Lucie M. C. R. Guibault & P. Bernt Hugenholtz eds., 2006).

153. Balkin, *supra* note 144, at 3–4 (explaining that an individual's ability to participate in the production and distribution of culture promotes freedom of speech).

154. *See* Jenny Kassan & Janelle Orsi, *The Legal Landscape of the Sharing Economy,* 27 J. ENV'T L. & LITIG. 1, 4–5 (2012).

accessibility of information about businesses, services, or commodities. This kind of information can help assess reputations and facilitate beneficial transactions, thereby minimizing inefficient services and streamlining markets.[155]

Third, dissemination of content promotes innovation, which is cumulative by nature, with most inventors building upon the work of their predecessors. Knowledge produces ideas that can be combined and recombined over and over again.[156] The free flow of information thus fosters the discussion of valuable ideas, promoting incremental innovation and enriching culture.

Fourth, dissemination of content allows the condemnation of immoral behavior, promotes desirable social norms and expands the scope of traditional shaming.[157] Today, "we live in a virtual 'global village' and events occurring across the world simultaneously occur in our living rooms."[158] This allows the public to express disapproval of individuals who violate social norms, empowering people to implement shaming sanctions and enforce social norms, which would otherwise go unenforced.[159] When the content

---

155. On the benefits of information flow between customers in promoting transparency in digital markets and improving various aspects of the standard form contract, see Shmuel I. Becher & Tal Z. Zarsky, *E-Contract Doctrine 2.0: Standard in the Age of Online User Participation*, 14 MICH. TELECOMM. & TECH. L. REV. 303, 314 (2008) ("Consumers advised of biased terms might refrain from contracting with specific vendors, should such contracting lead to inefficient outcomes. This information flow would stop vendors from including biased and unfair provisions in their SFCs to begin with, to avoid the loss of consumers."); Shmuel I. Becher & Tal Z. Zarsky, *Online Consumer Contracts: No One Reads but Does Anyone Care*, 12 JERUSALEM REV. L. STUDS. 105, 109 (2015) (giving an example of information on mandatory arbitration clauses in online terms that "further circulated and, as a result, the firm's reputation seemed to be coming under attack").

156. *See* ERIK BRYNJOLFSSON & ANDREW MCAFEE, THE SECOND MACHINE AGE: WORK, PROGRESS, AND PROSPERITY IN A TIME OF BRILLIANT TECHNOLOGIES 82 (2014) ("[K]nowledge itself increases over time as previous seed ideas are recombined into new ones. This is an innovative-as-building-block view of the world . . . ."). There is a virtually infinite number of potentially valuable reconfigurations of existing pieces of knowledge. On the cumulative nature of most innovation, see *White v. Samsung Electronics America, Inc.*, 989 F.2d 1512, 1515 (9th Cir. 1993): "All creators draw in part on the work of those who came before, referring to it, building on it, poking fun at it, we call this creativity, not piracy."

157. *See* Dan M. Kahan, *What Do Alternative Sanctions Mean*, 63 U. CHI. L. REV. 591, 611 (1996) (referring to the development of the shaming sanction in its early days and analyzing social condemnation through shaming as a valid form of criminal sanction). It is worth noting that several years later, Kahan acknowledged the shortcomings of shaming. *See* Dan M. Kahan, *What's Really Wrong with Shaming Sanctions*, 84 TEX. L. REV. 2075 (2005). Accordingly, everyone has different values and there remains ambiguity regarding the scope of immoral behavior that should be punished by shaming.

158. Lauren M. Goldman, *Trending Now: The Use of Social Media Websites in Public Shaming Punishments*, 52 AM. CRIM. L. REV. 415, 443 n.237 (quoting Deni Smith Garcia, *Three Worlds Collide: A Novel Approach to the Law, Literature, and Psychology of Shame*, 6 TEX. WESLEYAN L. REV. 105, 111 (1999)).

159. *See* Lavi, *supra* note 32, at 2601–02; Goldman, *supra* note 158, at 450 (2015) ("[T]he inclusion of online social media websites in public shaming sanctions may prove to be an effective form of punishment that takes into account the societal conditions that exist today."). An example of condemnation by shaming is a video recording a person's ugly behav-

disseminated is true and the dissemination does not violate the law,[160] sharing has many virtues[161] that may outweigh its vices.[162]

The conclusion may be different when the content disseminated includes falsehoods, defamation, fake news, or other information that causes harm without legitimate purpose.[163] In such cases, harmful information can spread rapidly, causing severe, unjustified reputational harm and even infringing on public interest. This can occur even when the original expression published is true if, during the process of dissemination, users add falsehoods to the original statement or cite the information and repeat it out of its original context.[164] Thus, even if the original expression reflects the truth, the process of dissemination can change the way the information is interpreted and increase misinformation, disinformation, and defamation.[165]

Individuals generally ascribe greater credibility to negative content than to other types of content.[166] This result is due to "negativity bias" and the strength of bad events, information, or feedback

---

ior. As the video is shared via social networks and goes viral, that individual is punished by being shamed. *See also, e.g.*, JON RONSON, SO YOU'VE BEEN PUBLICLY SHAMED ch. 1 (2015) (describing shaming that led to removal of a fake Twitter profile which hijacked a person's identity); Amit Cotler, *The Ugliest Kind of Israeli: Passengers Hurl Abuse at Flight Attendant*, YNET ISRAEL NEWS (Feb. 22, 2015), https://www.ynetnews.com/articles/0,7340,L-4629380,00.html [https://perma.cc/B8XY-RTN7].

160. Dissemination of certain types of content is illegal. For example, many states, and even other countries, criminalize the dissemination of nonconsensual pornography. *See* Mary Anne Frank, *Revenge Porn Reform: A View from the Front Lines*, 69 FLA. L. REV. 1251, 1256 (2017) (referring to state laws and the process of federal legislation in the United States); James Vincent, *Sharing Revenge Porn in the UK Now Carries a Two-Year Jail Sentence*, THE VERGE (Apr. 13, 2015), http://www.theverge.com/2015/4/13/8398691/revenge-porn-laws-uk-jail-time.

161. *See* DANIEL J. SOLOVE, THE FUTURE OF REPUTATION: GOSSIP, RUMOR AND PRIVACY ON THE INTERNET 92–94 (2007). In a world of increasingly rude and uncivil behavior, shaming helps society maintain a norm of civility and etiquette. Online shaming also gives people the chance to fight back, voice their disapproval of inappropriate behavior and even poor customer service. This kind of shaming allows the little guy to fight back against big corporations and also provides valuable information to help us assess another's reputation. Moreover, without the sanction of shaming, people would be able to continue rude and wrongful behavior without repercussion.

162. It should be noted that although Internet shaming has many benefits, it can also lead to some serious problems. For example, internet shaming is hard to control, can be disproportional, lacks due process and may lead to bullying. *See* SOLOVE, *supra* note 161, at 94–98; Lavi, *supra* note 32, at 2621.

163. *See generally* Danielle K. Citron, *Sexual Privacy*, 128 YALE L.J. 1870, 1908–24, 1929 (2019) (discussing the dissemination of intimate photos without consent and explains that current legal practices are ill equipped to accommodate the problem of abuses of digital dissemination).

164. Even dissemination of true information can develop into defamation when the disseminators combine new additions and headlines to the post and remove it from its original context. *See* Lavi, *supra* note 25, at 199–200; Lavi, *supra* note 32, at 2607 n.30.

165. MINA, *supra* note 99, at 148–50 (explaining that shared posts that spread norms and values (memes), can develop different interpretations and narratives and increase disinformation).

166. Lavi *supra* note 25, at 152.

compared to good events.[167] As a result, negative ideas receive more weight than other types of expression. Moreover, falsehoods tend to provoke and activate emotions and thereby be disseminated "farther, faster, deeper, and broader" than true statements.[168] Therefore, negative, defamatory content is likely to outweigh efforts by other social media users to counter it and continue to spread rapidly, ruining reputations.[169]

Alongside reputational harm, dissemination of defamatory remarks infringes on the same values that sharing strives to promote: freedom of expression, efficiency, innovation, and enforcement of norms.

First, the dissemination of falsehoods infringes upon the victim's free speech.[170] Due to the "negativity bias"[171] and "the weight ascribed to repeated content,"[172] diffusion of negative falsehoods "may lead to self-exclusion"[173] and "deny victims the ability to engage with others as equals, which might suppress a free public debate,"[174] harm the victim's autonomy, and hinder the free market of ideas and public participation.[175] Moreover, due to the technological and social context that allows the spread of falsehoods at the click of a button and bypasses deliberative thinking, dissemination may infringe on the disseminator's autonomy and free will.

---

167. *Id.*; Roy Baumeister, Ellen Bartslavsky, Catrin Finkenauer & Kathleen D. Vohs, *Bad Is Stronger than Good*, 5 REV. GEN. PSYCH. 323, 346 (2001); *see also* Elizabeth A. Kensinger, *Negative Emotion Enhances Memory Accuracy: Behavioral and Neuroimaging Evidence*, 16 CURRENT DIRECTIONS PSYCH. SCI. 213, 213 (2007) ("[N]egative emotion enhances not only the subjective vividness of a memory but also the likelihood of remembering some (but not all) event details.").

168. Vosoughi et al., *supra* note 15, at 1147; *see also* Jonah Berger & Katherine L. Milkman, *What Makes Online Content Viral?*, 49 J. MKTG. RSCH. 192, 197 (2012).

169. Lavi *supra* note 25, at 152.

170. *See* Jack M. Balkin, *How to Regulate (and Not Regulate) Social Media*, KNIGHT FIRST AMEND. INST. COLUM. U. (Mar. 25, 2020), https://knightcolumbia.org/content/how-to-regulate-and-not-regulate-social-media (explaining that the values that free speech is designed to serve are at risk).

171. *See* Baumeister et al., *supra* note 167, at 346.

172. *See* DIFONZO & BORDIA, *supra* note 47, at 225 (discussing the weight ascribed to repeated hearsay).

173. Lavi, *supra* note 25, at 182.

174. Lavi, *supra* note 53, at 53; *see also* Balkin, *supra* note 170 ("[A]ntagonistic sources of information do not serve the values of free expression when people don't trust anyone and professional norms dissolve."); CITRON, *supra* note 46, at 47–49 (referring to the potential of falsehoods to exclude individuals from public debate); Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 420 (2017) ("Individuals have difficulty expressing themselves in the face of online assaults."); Citron, *supra* note 115, at 1083. ("Online falsehoods, privacy invasions, and threats imperil targeted individuals' life opportunities, including their ability to express themselves.").

175. *See* Jeremy K. Kessler & David E. Pozen, *The Search for an Egalitarian First Amendment*, 118 COLUM. L. REV. 1953, 1994 (2018) ("[A]rguments involving *speech on both sides* focus on the degree to which one party's expressive activity compromises the ability of other private parties to exercise their own First Amendment rights.").

This is because the technological context might manipulate individuals to disseminate falsehoods without consideration, which they might regret at a later stage.[176]

In addition, spreading falsehoods within online social networks can distort the marketplace of ideas and the public interests of truth and democracy. Due to the technological context that allows sharing at the click of a button and the social context that facilitates information and reputation cascades, negative falsehoods can spread widely and users are more likely to perceive them as credible. The wide dissemination of falsehoods, defamation, and fake news erodes the truth and distorts the marketplace of ideas.[177] The socio-technological context of dissemination might not allow equal access to both "wise" and "unwise" ideas.[178] Falsehoods will spread farther than the truth, leading to an erosion of democracy and infringement of public interest.[179]

Second, beyond the infringement of free speech, spreading defamation and negative falsehoods about individuals injures their reputation, which is the basis for inclusion in market transactions. Thus, it may lead recipients to mistakenly avoid efficient transactions with individuals due to unchecked false information on them. As individuals are free to form contracts, can choose the people they contract with, and can avoid contracting with others, they are likely to avoid forming contracts with people who have had negative information disseminated about them, even if the information is false.[180] Third, the unjustified exclusion of individuals from markets may also hinder the development of new products and services. Fourth, due to negativity bias and the power of repetition,

---

176. *See* Alessandro Acquisti, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Gregory Norcie & Yang Wang, *I Regretted the Minute I Pressed Share: A Qualitative Study of Regrets on Facebook*, PROC. A.C.M. SYMP. ON USABLE PRIVACY & SECURITY, July 20-22, 2011, at § 4.3; Yang Wang, Pedro Giovanni Leon, Xiaoxuan Chen & Saranga Komanduri, *From Facebook Regrets to Facebook Privacy Nudges*, 74 OHIO ST. L.J. 1307, 1320–23 (2013).

177. FRANKS, *supra* note 115, at 119 ("[E]ven if people had strong preferences for the truth, there is no reason for confidence that the marketplace would help them discover it."); Mary Anne Franks & Ari Ezra Waldman, *Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions*, 78 MD. L. REV. 892, 894 (2019) ("[D]eliberately deceptive speech undermines, not enhances, the pursuit of truth."); *see also* Jonathan D. Varat, *Truth, Courage, and Other Human Dispositions: Reflections on Falsehoods and the First Amendment*, 71 OKLA. L. REV. 35, 48–49 (2018); Cass R. Sunstein, *Believing False Rumors*, *in* THE OFFENSIVE INTERNET: PRIVACY, SPEECH AND REPUTATION 91, 102 (Saul Levmore & Martha C. Nusbaum eds., 2010).

178. *Cf.* Dan Laidman, *When the Slander Is the Story: The Neutral Reportage Privilege in Theory and Practice*, 17 UCLA ENT. L. REV. 74, 99 (2010); Philip M. Napoli, *What if More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble*, 70 FED. COMM. L.J. 55, 69 (2018); Sunstein, *supra* note 8, at 406 ("[T]he marketplace of ideas can fail, ensuring that false statements will spread and become entrenched.").

179. *See* Hasen, *supra* note 12, at 544; Sunstein, *supra* note 8, at 394 ("[I]f people spread false statements—most obviously about public officials and institutions—democracy itself will suffer.")

180. *See* RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 266–67 (8th ed. 2011).

dissemination of defamation validates falsehoods and develops undesirable social perceptions.

In sum, dissemination of content has many benefits. It promotes values of free expression, efficiency, innovation, and the enforcement of desirable social norms. Yet, when the content disseminated is false and negative, the same values can be infringed and severe harm can be inflicted. The law should not stop the flow of information and it must avoid a disproportionate "chilling effect" on speech. Solutions, however, should be developed to mitigate harm caused by the dissemination of falsehood. Protecting individual reputations and the public interest of access to the truth, without curbing free speech, is one of the main challenges internet regulation faces today.

II. SECONDARY LIABILITY OF INTERMEDIARIES

How does the law deal with dissemination of falsehoods on social network platforms? This Part provides a comparative overview focusing on intermediary liability for defamation.[181] It continues with a critical review of international policy models governing secondary liability of online intermediaries that facilitate the harmful exchange of information.

A. *United States*

In the United States, freedom of speech has greater protection than in other Western democracies: there is a presumption against restrictions on speech.[182] Section 230 of the CDA provides one of

---

181. This Part will focus on the intermediary and not on the liability of the publisher of the harmful expression. The liability of the publisher can only be helpful in limited contexts, and cannot prevent the further spread of information, as the law immunizes online republishers. The Article will not address the potential liability of individuals who share information since it is difficult and even impractical to hold them responsible and receive full compensation for the aggregated reputational damage. Litigation can be cumbersome since none of the repolishes that shared content is solely responsible for the damage and there is a large number of defendants. For further information on this aspect of dissemination, see Perry, *supra* note 26. This Part will also not address specific election campaign finance laws that apply to publishers. For further information on this topic, see Hasen, *supra* note 12, at 554–63.

182. Evelyn Douek, *Governing Online Speech From "Posts-as-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. (forthcoming 2021) (manuscript at 11–12); Oreste Pollicino & Marco Bassini, *Free Speech, Defamation and the Limits to Freedom of Expression in the EU: A Comparative Analysis*, *in* RESEARCH HANDBOOK ON EU INTERNET LAW 513–28 (Andrej Savin & Jan Trzaskowski eds., 2014) (demonstrating that in the United States, there are stronger protections on the freedom of speech than in the EU, and that the different balance between values is even more prominent online). *But see* FRANKS, *supra* note 115 (arguing that legislators,

the most important protections of freedom of expression in the United States in the digital age.[183] This federal legislation represents the mindset of internet exceptionalism, differentiating the internet from the media before.[184] It generally blocks lawsuits against online intermediaries. Section 230(c)(1), under the subsection header "Protection for 'Good Samaritan' Blocking and Screening of Offensive Material," directs that "[n]o 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."[185] "Congress declared that online service providers could never be treated as publishers for material they did not develop."[186]

By declaring that online intermediaries cannot be treated as publishers of content authored by others, the aim of Congress was to promote self-regulation, freedom of expression, and support the rise of lively internet enterprises.[187] Under § 230(c)(1), online service providers, including website operators, are immune from primary and secondary liability for a wide variety of claims.[188] This is because an individual user who is considered "another information content provider" published the original defamatory content.[189] Intermediaries are not held responsible for that user's conduct or for republishing content. This applies even when the intermediary it-

---

courts and civil rights organizations have interpreted the First Amendment selectively, just like religious fundamentalists, and in fact infringe on the rights of minorities and the underprivileged to free speech, shifting even more power from vulnerable populations to powerful ones).

183. Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2313 (2014); *see also* Eric Goldman, *Why Section 230 Is Better than the First Amendment*, 95 NOTRE DAME L. REV. ONLINE 33 (2019).

184. KOSSEFF, *supra* note 20, at 78.

185. Communication Decency Act, 47 U.S.C. § 230(c)(1).

186. Anupam Chander, *How Law Made Silicon Valley*, 63 EMORY L.J. 639, 651 (2014). The Protection for Good Samaritan subsection aims to promote self-regulation and encourage intermediaries to screen offensive materials without bearing liability. *See* Citron & Wittes, *supra* note 174, at 403; Jonathan Zittrain, *A History of Online Gatekeeping*, 19 HARV. J.L. & TECH. 253, 262–63 (2006).

187. For further information on the history and objectives of § 230, see JEFF KOSSEFF, *supra* note 20, at 11–35 (2019); Chander, *supra* note 186, at 651–52; and Lavi, *supra* note 32, at 2636.

188. *See, e.g.*, Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997) ("By its plain language, § 230 creates a federal immunity to *any* cause of action that would make service providers liable for information originating with a third-party user of the service" (emphasis added).); Blumenthal v. Drudge, 992 F. Supp. 44, 51–53 (D.D.C. 1998) (finding that intermediaries are immune even if they pay third parties to write columns on their platforms, which contain defamatory speech); Ben Ezra, Weinstein, & Co. v. Am. Online Inc., 206 F. 3d 980, 984–86 (10th Cir. 2000) (holding that intermediaries are immune even when they provide access to information from third parties and this information is erroneous).

189. 47 U.S.C. § 230(c)(1).

self republishes the user's content.[190] Courts have interpreted § 230 broadly and have also shielded intermediaries from liability as distributers of content.[191] As a result, this section has "repeatedly shielded web enterprises from lawsuits in a plethora of cases" when they failed to remove harmful content, when they operated editorial discretion and discriminated content, and even when they performed more active roles in dissemination of content.[192] It should be noted that this immunity is gradually eroding. First, intermediaries are only immune with regard to information that is "provided by another content provider."[193] If a plaintiff can demonstrate that the intermediary provided the content, the intermediary will not benefit from § 230 immunity. Second, even if a third party is held accountable for creating the content, § 230 only prevents the court from treating the platform as "publisher or speaker." If a plaintiff can demonstrate that the lawsuit stemmed from an action by the defendant that was not publishing or speaking, the court should find that § 230 does not block the lawsuit.[194]

---

190. Lavi, *supra* note 25; *see also, e.g.*, Batzel v. Smith, 333 F.3d 1018, 1020, 1034 (9th Cir. 2003) (holding that a moderator of a listserv and operator of a website who posted a defamatory e-mail authored by a third party may be exempt from liability if the material is "provided" by someone else); Roca Labs, Inc. v. Consumer Opinion Corp., 140 F. Supp. 3d 1311, 1318–20 (M.D. Fla. 2015). It is worth noting that § 230 also exempts internet users who share content published by others. *See* Barrett v. Rosenthal, 146 P.3d 510, 525, 528–29 (Cal. 2006) ("[C]ongressional purpose of fostering free speech on the Internet supports the extension of section 230 immunity to active individual 'users.' ").

191. Cecilia Ziniti, *The Optimal Liability System for Online Service Providers: How* Zeran v. America *Online Got It Right and Web 2.0 Proves It*, BERKELEY TECH. L.J. 583, 585–87 (2008); *cf.* 47 U.S.C. § 230(b)(1)–(2).

192. Lavi, *supra* note 54, at 867–70; *see also Zeran*, 129 F.3d at 330 ("By its plain language, § 230 creates a federal immunity to any cause of action that would make service providers liable for information originating with a third-party user of the service."); Nemet Chevrolet, Ltd. v. Consumeraffairs.com, Inc., 591 F.3d 250, 254 n.4 (4th Cir. 2009); Giordano v. Romeo 76 So. 3d 1100, 1101–02 (Fla. Dist. Ct. App. 2011); Caraccioli v. Facebook, Inc., 167 F. Supp. 3d 1056, 1064–66 (N.D. Cal. 2016) (holding that immunity applies even when the intermediary knew of the defamatory content and did not remove it), *aff'd*, No. 16-15610, 2017 WL 2445063 (9th Cir. June 6, 2017); Herrick v. Grindr LLC, No. 18-396, 2019 WL 1384092 (2d Cir. Mar. 27, 2019); Prager Univ. v. Google LLC, 951 F.3d 991, 995 (9th Cir. 2020) (holding that activity of intermediaries to restrict materials is covered by § 230's immunity as intermediaries are not state actors and are not subjected to the First Amendment); Fyk v. Facebook, Inc., No. 19-16232, 2020 WL 3124258, at *1 (9th Cir. June 12, 2020) (noting that § 230 protects intermediaries' editorial discretion to moderate content); *Batzel*, 333 F.3d at 1030–33 (holding that immunity applies even when an operator of a listserv repeated users' content in a listserv).

193. 47 U.S.C. § 230(c)(1) (referring to immunity for "any information provided by another information content provider"); KOSSEFF, *supra* note 20, at 166.

194. KOSSEFF, *supra* note 20, at 166; Lavi, *supra* note 32, 2659 n.251; *see also* Harrington v. Airbnb, Inc., 348 F. Supp. 3d 1085 (D. Or. 2018) (noting that "because a local regulation did not require the Platforms to monitor third-party content" or to remove it, it does not treat them as publishers, and thus falls outside the preemptive scope of § 230); Eric Goldman, *Racial Discrimination Lawsuit Against Airbnb Has the Potential to Change Online Marketplaces*–Harrington v. Airbnb, TECH. & MKTG. LAWBLOG (Nov. 2, 2018) (explaining that although the case does not discuss § 230, it offers a roadmap around it); Bolger v Amazon.com, Inc., 53 Cal. App. 5th 431 (2020) (holding that Amazon is not immune from liability to market-

Section 230, however, recently sustained an attack regarding immunity for user generated content. Recently, Twitter added a fact checking label to tweets of the 45th President of the United States, Donald Trump, stating that viewers could "get the facts" by clicking on the addendum.[195] Following this labeling, President Donald Trump attempted to curb the online platform's protection for "Good Samaritans."

On May 28, 2020, Trump issued an executive order on "Preventing Online Censorship" pertaining to online platforms.[196] After a policy statement on the need to "seek transparency and accountability from online platforms, and . . . preserve the integrity and openness of American discourse and freedom of expression,"[197] the order outlines a narrow interpretation to § 230. It clouds the legal landscape for content moderation decisions, explaining that § 230(c)(2) applies only to "good faith" moderation decisions.[198] It, thus, allows stripping the shield from moderation decisions that the government does not see as moderation in "good faith." The order further directs "all executive departments and agencies" to "ensure that their application of [S]ection 230(c) properly reflects the narrow purpose of the section and take all appropriate actions in this regard."[199]

In addition, the order directs each executive department and agency to review media advertising spent on online platforms and restrict platforms' receipt of advertising dollars.[200] "The Department of Justice shall review in the viewpoint-based speech restrictions imposed by each online platform . . . and assess whether any online platforms are problematic vehicles for government speech due to viewpoint discrimination, deception to consumers, or other bad practices."[201] The order further provides that the White House "will submit" reports of purported "online censor-

---

place items); Huon v. Denton, 841 F.3d 733, 736–37 (7th Cir. 2016) (alleging that the intermediary and its employees wrote the defamatory statements). For criticism of extending § 230 liability to other activities, see Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Speech Reform*, 2020 U. CHI. LEGAL F. 45, 51–52.

195. *See* Makena Kelly, *Twitter Labels Trump Tweets as 'Potentially Misleading' for the First Time*, THE VERGE (May 26, 2020, 6:04 PM), https://www.theverge.com/2020/5/26/21271207/twitter-donald-trump-fact-check-mail-in-voting-coronavirus-pandemic-california.

196. Exec. Order No. 13,925, 85 Fed. Reg. 34,079 (June 2, 2020).

197. *Id.* § 1.

198. *See id.* § 2 ("[U]nder the law, this provision is not distorted to provide liability protection for online platforms that—far from acting in "good faith" to remove objectionable content — instead engage in deceptive or pretextual actions (often contrary to their stated terms of service) to stifle viewpoints with which they disagree.").

199. *Id.* § 2(b).

200. *Id.* § 3.

201. *Id.* § 3(c).

ship" received through its "Tech Bias Reporting Tool" to the Department of Justice and the Federal Trade Commission (FTC).[202] The latter can "consider taking action" under applicable law, including under Section 5 of the FTC Act,[203] which makes unfair methods of competition unlawful.[204]

Legal experts agree that the order is without legal foundation, unenforceable, and without legal impact.[205] Recently the Center for Democracy & Technology filed a lawsuit against the executive order to invalidate it.[206] In addition, the Court of the Northern District of New York ruled that the executive order precluded a private right of action even if defendants in that case arbitrarily removed the plaintiff's account or prevented him from creating a new account.[207] It is therefore likely that the immunity provided under § 230 will remain strong where platforms host harmful content created by third parties, moderate content, and allow users to share it.

In addition to the order, recent legislative bills strive to narrow § 230's immunity.[208] The shadow of the order and the legislative bills, however, might impair how intermediaries apply their First Amendment rights to moderate content and lead all platforms to chill more protected speech.[209]

B. *A Comparative Perspective*

In Europe, the scope of intermediary liability is much broader than in the United States, and the balance between freedom of speech and protection of reputation is very different.[210] The E-

---

202. *Id.* § 4(b).

203. Federal Trade Commission Act, 15 U.S.C § 45.

204. Exec. Order, *supra* note 196, at § 4(c).

205. *See, e.g.*, Eric Goldman, *Trump's "Preventing Online Censorship" Executive Order Is Pro-Censorship Political Theater*, TECH. & MKTG. L. BLOG (May 29, 2020), https://bit.ly/2B33vSk; Jan Wolfe, *Trump's Order Taking Aim at Twitter Is 'Bluster': Legal Experts*, REUTERS (May 28, 2020, 2:17 PM), https://www.reuters.com/article/us-twitter-trump-executive-order-analysis/trumps-order-taking-aim-at-twitter-is-bluster-legal-experts-idUSKBN234361 [https://perma.cc/4CRN-759Z].

206. Complaint, Ctr. For Tech. & Democracy v. Trump, No. 1:20-cv-01456 (D.D.C. Dec. 11, 2020), 2020 WL 2858041.

207. Gomez v. Zuckerburg, No. 5:20-cv-00633-TJM-TWD, 2020 U.S. Dist. LEXIS 130989 (N.D.N.Y. July 23, 2020); *see also* Eugene Volokh, *No Claim Against Facebook Based on President's Social Media Executive Order*, VOLOKH CONSPIRACY (July 31, 2020), https://reason.com/volokh/2020/07/31/no-claim-against-facebook-based-on-presidents-social-media-executive-order/ [https://perma.cc/K3DP-P6S4].

208. *See* S. 4534, 116th Cong. (2019); S. 4066, 116th Cong. (2020).

209. *See* Complaint, *supra* note 206, at ¶ 45.

210. *See* Neil Richards & Woodrow Hartzog, *Privacy's Constitutional Moment*, 61 B.C. L. Rev. 1687, 1729–30 (2020) ("In Europe, free expression is safeguarded by Article 10 of the European Convention and Article 11 of the EU Charter. Like other European fundamental

Commerce Directive dictates the framework for intermediary liability. According to Article 14, intermediaries that host content are subject to a 'notice-and-takedown' regime that obligates them to remove illegal content in order to avoid potential liability. [211] "This knowledge-based safe haven protects intermediaries whose role is 'merely technical, automatic and passive,' but does not shield intermediaries that play an active role in hosting the content."[212]

The Directive does not prevent Member States from establishing specific requirements nor does it affect orders by national authorities in accordance with national legislation.[213] Member States can impose duties of care on intermediaries through national legislation, requiring them to undertake reasonable efforts to detect and prevent certain types of illegal activity.[214] Thus, for example, in the fall of 2017, the German government drafted the Network Enforcement Act (NetzDG) for targeting hate speech and fake news.[215] The Act applies to criminally offensive speech as defined in the German Penal Code, including defamation.[216] It stipulates a

---

rights, these provisions are subject to proportionality analysis – where they conflict with another fundamental right such as the right to privacy or to data protection, courts must balance the rights on an equal footing. By contrast, in the United States, the fundamental right of free expression protected by the First Amendment is not subject to proportionality analysis . . . ."). *See generally* Pollicino & Bassini, *supra* note 182.

211. Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market, 2000 O.J. (L 178) 1, Art. 14(1) [hereinafter E-Commerce Directive] ("Where an information society service is provided that consists of the storage of information provided by a recipient of the service, Member States shall ensure that the service provider is not liable for the information stored at the request of a recipient of the service, on condition that: (a) the provider does not have actual knowledge of illegal activity or information and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent; or (b) the provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information."). For expansion on the E-Commerce Directive, see Lavi, *supra* note 54, at 870–71.

212. Lavi, *supra* note 53, at 46 (citing Joined Cases C-236 & C-238/08, Google France, S.A.R.L. & Google Inc. v. Louis Vuitton Malletier SA, 2010 E.C.R. I-2417, I-2513 ("[T]o establish whether the liability of a referencing service provider may be limited under Article 14 of Directive 2000/31, it is necessary to examine whether the role played by that service provider is neutral, in the sense that its conduct is merely technical, automatic and passive, pointing to a lack of knowledge or control of the data which it stores."); and Corey Omer, *Intermediary Liability for Harmful Speech: Lessons from Abroad*, 28 HARV. J.L. & TECH. 289, 313 (2014)).

213. E-Commerce Directive, *supra* note 211, at Recital 47.

214. *Id.* at Recital 48.

215. Netzwerkdurchsetzungsgesetz [NetzDG] [Network Enforcement Act], Oct. 1, 2017, at § 3(2)(4) (Ger.), *translation at* https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf?__blob=publicationFile&v=2 [https://perma.cc/PH7B-6FYQ].

216. Wolfgang Schulz, *Regulating Intermediaries to Protect Privacy Online – the Case of the German NetzDG*, in PERSONALITY AND DATA PROTECTION RIGHTS ON THE INTERNET (Marion Albers & Ingo Sarlet eds.) (forthcoming) (manuscript at 5) (citing GERMAN PENAL CODE, §§ 185–189, *translation at* http://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html [https://perma.cc/MQ89-WSQA]).

differential timeframe for intermediaries to remove harmful content.[217] Intermediaries "have to make sure that they delete content that appears . . . evidently unlawful within 24 hours" of the filing of a complaint.[218] "When content is not evidently unlawful," intermediaries have to remove it within seven days.[219] The "[r]eview period may exceed 7 days when more time is required for the decision-making[, in order to minimize] 'over-blocking.'"[220] Failure to comply with the law can lead to a fine of up to five million Euros.[221]

The Directive is somewhat obsolete in that its classification may no longer fit to each and every role intermediaries perform today. Many intermediaries may not be classified as "hosts" since courts interpret Article 14 narrowly.[222] Outside the scope of the E-Commerce Directive intermediaries' liability can be broad.[223] The case of *Delfi* is one good example. In this case, "the Estonian Supreme Court found the popular Delfi news website liable for defamatory statements about a famous Estonian business executive."[224] Although Delfi followed a proper "notice-and-takedown" regime and complied with the Directive,[225] the Directive's safe haven was not applied because "by allowing comments from unregistered and anonymous users, the site is liable as a publisher."[226] Thus, it was not considered a "host."

Delfi then appealed to the European Court of Human Rights (ECHR),[227] claiming a violation of freedom of expression.[228] The

---

217. *Id.*

218. *Id.*

219. *Id.*

220. *Id.*; *see also* NetzDG, *supra* note 215, at § 3(2)(3). For further information and criticism, see BENKLER ET AL., *supra* note 2, at 362–63; and Schulz, *supra* note 216.

221. Schulz, *supra* note 216, at 6.

222. *See* Lavi, *supra* note 54, at 871; Ronen Perry & Tal Zarsky, *Liability for Online Anonymous Speech: Comparative and Economic Analyses*, 5 J. EUR. TORT L. 205 (2014); Peggy Valcke & Marieke Lenaerts, *Who's Author, Editor and Publisher in UGC Content? Applying Traditional Media Concepts to UGC Providers*, 24 INT'L REV. L. COMPUTS. & TECH. 119, 126 (2010).

223. Lavi, *supra* note 54, at 871–74.

224. *Id.* at 871–72; *see also* Delfi AS v. Estonia, 2015-II Eur. Ct. H.R. 319, 344 (June 16, 2015).

225. Perry & Zarsky, *supra* note 222, at 221.

226. Lavi, *supra* note 54, at 872; *see also* Delfi AS v. Estonia, App. No. 64569/09, ¶28 (Oct. 10, 2013), https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-126635%22]} (noting that, in reviewing the decision of the Estonian Court, "the Supreme Court considered that in the present case both the applicant company and the authors of the comments were to be considered publishers of the comments").

227. The ECHR "is charged with supervising the enforcement of the European Convention for the Protection of Human Rights . . . drawn up by the Council of Europe. . . . Individuals who believe their human rights have been violated and who are unable to remedy their claim through their national legal system may petition the ECHR to hear the case and render a verdict." John G. Merrills, *European Court of Human Rights*, BRITANNICA, https://www.britannica.com/topic/European-Court-of-Human-Rights.

228. *Delfi*, 2015-II Eur. Ct. H.R. at 327 ("The applicant company alleged that its freedom of expression had been violated.").

First Section of the ECHR did not accept Delfi's claim.[229] It upheld the Estonian Court's ruling, finding that the ruling was in line with Article 10 of the European Convention on Human Rights,[230] as a proportional interference with the freedom of expression. The Grand Chamber confirmed the decision.[231] That decision has created confusion regarding what distinguishes online "publishers" from mere intermediaries.[232]

A narrow interpretation of the E-Commerce Directive was also recently applied by the European Court of Justice (ECJ).[233] In *Glawischnig-Piesczek v. Facebook Ireland Ltd.,* a member of the National Council of Austria sued Facebook Ireland in the Austrian courts, seeking an order for Facebook to remove an allegedly defamatory comment about her and equivalent comments.[234] Although the E-Commerce Directive does not stipulate a general monitoring obligation, the ECJ held that intermediaries such as Facebook are not protected by EU Law from an order to remove content that is identical, and even potentially similar to, content previously declared unlawful.[235]

At present, the extent of the "notice-and-takedown" provisions is unclear and it appears that the E-commerce Directive's safe haven is eroding.[236] A narrow interpretation of "hosting" enables courts to hold intermediaries accountable for negligence in preventing third-party harm, despite removing defamatory content from their platforms upon knowledge.[237] In addition, requiring the removal of similar or equivalent content when that content has already been

---

229. *Delfi*, 2015-II Eur. Ct. H.R. 319.

230. *Id.* at 359, 366–67; *see also* Convention for the Protection of Human Rights and Fundamental Freedoms art. 10, Nov. 4, 1950, 213 U.N.T.S. 2889. The Court applied a narrow interpretation of intermediary technical functions. For further information, see Martin Husovec, *ECtHR Rules on Liability of ISPs as a Restriction of Freedom of Speech*, 9 J. INTELL. PROP. L. & PRAC. 108, 109 (2014)1

231. *Delfi*, 2015-II Eur. Ct. H.R. at 388; *see also* Lavi, *supra* note 54, at 872–73.

232. Perry & Zarsky, *supra* note 222, at 222; Lavi, *supra* note 54, at 872–73; Lavi, *supra* note 53, at 48.

233. *See generally Court of Justice in the European Union*, EUROPA.EU, https://europa.eu/european-union/about-eu/institutions-bodies/court-justice_en [https://perma.cc/S77K-GE66] (last visited Nov. 4, 2020).

234. Case C-18/18, Glawischnig-Piesczek v. Facebook Ir. Ltd., ECLI:EU:C:2019:821 (Oct. 3, 2019) (reviewing the decision of the Vienna Commercial Court).

235. Lavi, *supra* note 141, at 508; *see also Glawischnig-Piesczek*, ECLI:EU:C:2019:821. It should be noted that the obligation is to block access to the information worldwide and not only from within the EU domains of Facebook. For criticism that the obligation to remove similar and equivalent content can lead to over-blocking and have a chilling effect on free speech, see DAPHNE KELLER, STANFORD CENTER FOR INTERNET AND SOCIETY, DOLPHINS IN THE NET: INTERNET CONTENT FILTERS AND THE ADVOCATE GENERAL'S *GLAWISCHNIG-PIESCZEK V. FACEBOOK IRELAND* OPINION 18–19 (2019) (comparing false positives to dolphins in the net).

236. Lavi, *supra* note 54, at 871.

237. Delfi AS v. Estonia, 2015-II Eur. Ct. H.R. 319, 365–66 (June 16, 2015).

deemed unlawful imposes an obligation on intermediaries to use the technology available to them to monitor the platform.[238] This obligation is beyond the traditional knowledge-based safe haven outlined in the E-Commerce Directive.[239]

The interpretation of the European Data Protection Directive on the "right to be forgotten" also reflects an expansion of intermediary liability.[240] In *Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos*,[241] the ECJ supported what is known as the "right to be forgotten."[242] Specifically, the ECJ held that search engines, like Google, must remove search results that link to personal information, including defamatory content found on third party websites upon user request.[243]

"The ECJ reached this conclusion by broadly interpreting the term 'controller' in Article 2(d) of the Data Protection Directive,"[244] affirming that indexing personal data published on websites makes search engines data processors and controllers.[245] By classifying search engines as controllers, the court implies that "they are not neutral and passive enough to be eligible for the safe harbors' protection."[246] In a recent decision, the ECJ held that "the right to be forgotten" should apply to "search engine versions ac-

---

238. KELLER, *supra* note 235, at 11–12.

239. *See generally* E-Commerce Directive, *supra* note 212, at art. 14(1).

240. *See* Council Directive 95/46, 1995 O.J. (L 281) 31 (EC) [hereinafter Data Protection Directive].

241. Case C-131/12, ECLI:EU:C:2014:317 (May 13, 2014).

242. *See* Lavi, *supra* note 54, at 873 n.82 ("This right, now branded as the 'right to erasure,' was represented as one of the 'four pillars' of the new Regulation in the European Union. In October 2013, the European Parliament Committee on Civil Liberties, Justice, and Home Affairs considered and consolidated nearly four thousand proposed amendments to the Commission Proposal into a new proposal that was adopted by the Committee."). For an in-depth discussion, see Cooper Mitchell-Rekurt, *Search Engine Liability Under the Libe Data Regulation Proposal: Interpreting Third Party Responsibilities as Informed by Google Spain*, 45 GEO. J. INT'L L. 861 (2014); Abraham L. Newman, *What the "Right to be Forgotten" Means for Privacy in a Digital Age*, SCI. MAG., Jan. 30, 2015, at 507; and NEIL RICHARDS, INTELLECTUAL PRIVACY–RETHINKING DIGITAL LIBERTIES IN THE DIGITAL AGE 91 (2015).

243. Lavi, *supra* note 25, at 173; *see also* Miquel Peguera, *The Shaky Ground of the Right To Be Delisted*, 18 VAND. J. ENT. & TECH. L. 507, 549 (2016) (explaining that this ruling is confined to searches made using the name of an individual and asserts a right to delist the link—as opposed to a right to remove the information from the search engine's index altogether); Anupam Chander & Uyen P. Le, *Free Speech*, 100 IOWA L. REV. 501, 541 (2015) ("To comply with the judgment, Google offered EU citizens the ability to file data removal requests. Within 24 hours, the search engine received right to be forgotten requests from at least 24,000 individuals."); MEG LETA JONES, CTRL + Z: THE RIGHT TO BE FORGOTTEN 10, 41, 46 (2016).

244. Data Protection Directive, *supra* note 240, at art. 2(d) ("Article 2(d) 'controller' shall mean the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data.").

245. Lavi, *supra* note 54, at 874; *see also Google Spain*, ECLI:EU:C:2014:317 at ¶ 41; Lavi, *supra* note 32, at 2632.

246. Lavi, *supra* note 25, at 173; *see also* Peguera, *supra* note 243, at 544.

cessible in EU Member States, as opposed to all versions of its search engine worldwide."[247]

It should be noted that the Data Protection Directive was replaced by the General Data Protection Regulation (GDPR) in May 2018.[248] The GDPR includes a specific provision titled "Right to erasure ('right to be forgotten')"[249] that imposes data controller obligations to erase data.[250] In general, the GDPR imposes more specific obligations regarding information processing,[251] and the ECJ interprets intermediaries' obligations in this regard broadly.[252]

Other countries outside Europe outline different intermediary liability regimes.[253] Some jurisdictions have even passed anti-fake news laws that are beyond private law, addressing infringement of public interest.[254] For example, a new law in Singapore allows the *government* to order intermediaries to remove false statements.[255]

---

247. Harlan Grant Cohen & Monika Zalnieriute, *Google LLC v. Commission Nationale de l'Informatique et des Libertés (CNIL)*, 114 AM. J. INT'L L. 261 (2020); *see also* Case C-507/17, Google LLC v. Commission Nationale de L'informatique et des Libertés (CNIL), ECLI:EU:C:2019:772 (Sept. 24, 2019).

248. The GDPR subjects "controllers" to a broader right to erasure. Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46, art. 17, 2016 O.J. (L119) 33 [hereinafter GDPR]; *see also* Chris Jay Hoofnagle, Bart van der Sloot & Frederick Zuiderveen Bogesius, *The European Union General Data Protection Regulation: What It Is and What It Means*, 28 INFO. & COMM. TECH. L. 65, 90 (2019) ("Roughly summarized, a data subject has a right to erasure when he or she successfully exercises the right to object, when the personal data were unlawfully processed, should be erased because of a legal obligation, or are no longer necessary in relation to the processing purposes."); Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For,* 16 DUKE L. & TECH. REV. 18, 68–69 (2017).

249. GDPR art. 17; Lavi, *supra* note 32, at 2634.

250. GDPR art. 17; *see also* Lavi, *supra* note 32, at 2634; JONES, *supra* note 243, at 10 (explaining that Article 17 to the GDPR, titled "Right to erasure ('right to be forgotten,')" can impose obligations on controllers to delete information from the internet altogether).

251. *See* Michael L. Rustad & Thomas H. Koenig, *Towards a Global Data Privacy Standard*, 71 FLA. L. REV. 365 (2019).

252. *E.g.*, C-136/17, GC v. Commission Nationale de l'Informatique et des Libertés (CNIL), ECLI:EU:C: 2019:773 (Sept. 24, 2019).

253. For further information, see Lavi, *supra* note 25, at 174, which expands on intermediary liability in Canada.

254. *See The Rise of "Fake News" Laws Across South East Asia,* PUB. MEDIA ALLIANCE (Dec. 6, 2019), https://www.publicmediaalliance.org/the-rise-of-fake-news-laws-across-south-east-asia/ (overviewing Fake News Laws across South East Asia particularly on media freedom).

255. Protection from Online Falsehoods and Manipulation Bill, Parl. Bill No 10/2019, § 4 (2019), https://www.parliament.gov.sg/docs/default-source/default-document-library/protection-from-online-falsehoods-and-manipulation-bill10-2019.pdf (section four of the law refers to directions to internet intermediaries and providers of mass media services); *see also* Jason Luger, *Planetary Illiberalism and the Cybercity-state: In and Beyond Territory*, 8 TERRITORY, POL., GOVERNANCE 1 (2019); Niharika Mandhana & Phred Dvorak, *Ordered by Singapore, Facebook Posts a Correction,* WALL ST. J. (Nov. 30, 2019), https://www.wsj.com/articles/facebook-complies-with-order-under-singapore-fake-news-law-11575116149. Such laws were aimed at preserving political security, but governments can use them to prevent damage to the public's health that can occur as a result of believing fake news on Covid-19. *See* Ellie Bothwell, *Fake News Laws May 'Catch On' During Coronavirus*, TIMES HIGHER ED.

Yet, this regime poses a threat to free speech since it subjects the intermediary directly to the government.[256] Moreover, the absence of any definition of the term "false statement of fact" provides the government overly broad discretion.[257]

C. *Liability Regimes: A Critical View*

"Intermediary liability rests on the junction of a few areas of law. It balances constitutional rights and tort considerations,"[258] aiming to find the right balance between values. On the one hand, the digital ecosystem allows anyone to publish harmful statements and potentially infringe upon the victim's right to reputation and free speech. On the other hand, liability can lead to collateral censorship "when a (private) intermediary suppresses the speech of others in order to avoid liability" for such speech.[259] Imposing liability on the intermediary for false rumors may also constitute an infringement on the freedom to conduct a business.[260] In the United States, "an individual's right to conduct a business or pursue an occupation is a property right."[261] Yet, freedom to conduct a business is not absolute. Companies are subject to certain basic requirements and remain accountable for decisions that might infringe on individual rights.[262]

In addition to constitutional balances, "the technological context of intermediary liability involves considering the influence of liability on the path of innovation" and its repercussions on welfare maximization.[263] Courts should, therefore, not solely consider the harm to victims but also the benefits of an activity to third par-

---

(Apr. 6, 2020), www.timeshighereducation.com/news/fake-news-laws-may-catch-during-coronavirus.

256. *See* Tessa Wong, *Singapore Fake News Law Polices Chats and Online Platforms*, BBC NEWS (May 9, 2019), https://www.bbc.com/news/world-asia-48196985.

257. On the flaws of specific legal liability frameworks for "fake news" and the difficulty of defining "fake news," see MINA, *supra* note 99, at 126.

258. Lavi, *supra* note 53, at 49.

259. Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293, 295–96 (2011).

260. *See* Charter of Fundamental Rights of the European Union, Dec. 7, 2000, 2000 O.J. (C 364) art. 16 (expressing that the EU was founded on the universal values of dignity, solidarity, freedom, and equality).

261. Garrison v. Herbert J. Thomas Mem'l Hosp. Ass'n, 438 S.E.2d 6, 7 (W. Va. 1993); *see also* United States v. Santoni, 585 F.2d 667, 673 (4th Cir. 1978); United States v. Arena, 180 F.3d 380, 394 (2d Cir. 1999).

262. HARTZOG, *supra* note 120, at 121 ("Companies should generally have the freedom to design technologies how they please, so long as they stay within particular thresholds, satisfy certain basic requirements like security and accuracy, and remain accountable for deceptive, abusive, and dangerous design decisions.").

263. Lavi, *supra* note 53, at 49.

ties.[264] Their balance should include the overall costs and benefits to society as a whole.[265] Finding the right balance between these interests is the key to formulating a proper intermediary liability regime. It should be noted that the balance between values is subjective and differs between legal systems.

Different countries apply different policy models for intermediary liability, such as: overall immunity,[266] a safe haven provision ("notice-and-takedown"),[267] and negligence liability for failing to take reasonable precaution to prevent third-party harm.[268] These policy models are either over- or under-inclusive. This section analyzes the shortcomings of common liability regimes governing secondary intermediary liability and argues that such regimes fail to accommodate the challenges of dissemination of false rumors.

In a previous Article, I reached the conclusion that a "notice-and-takedown" safe haven regime is preferable to other regimes for regulating secondary liability of intermediaries on social network platforms.[269] This type of regime is a compromise.[270] Under such a regime, intermediaries are not required to block or filter content and they do not bear liability for harmful content they were not informed about.[271] Only intermediaries that fail to remove harmful content upon notice expose themselves to liability.[272] In light of the extensive harm false rumors and defamation may cause in social networks, this outcome is appropriate in comparison to an immunity regime.[273]

This analysis applies to liability for original speech. The vast multitude of possibilities to quickly share content on social networks undermines the efficiency of a "notice-and-takedown" regime, since false rumors and defamatory content are speedily replicated. A victim aspiring to remove defamatory remarks would need to send a complaint and indicate each and every virtual loca-

---

264. *Id.* at 56.
265. Lavi, *supra* note 141, at 536.
266. *See* 47 U.S.C. § 230 (U.S. model).
267. *See* 2000 O.J. (L 178) 13 (EU model).
268. *See* Delfi AS v. Estonia, 2015-II Eur. Ct. H.R. 319, 344 (June 16, 2015) (Estonian Model).
269. *See* Lavi, *supra* note 52, at 930–31.
270. *See* Celia Ziniti, *The Optimal Liability System for Online Service Providers: How* Zeran v. America Online *Got It Right and Web 2.0 Proves It*, 23 BERKELEY TECH. L.J. 583, 604–07 (2008) (explaining that "notice-and-takedown" regimes can essentially lead to the removal of any content in response to any complaint).
271. Lavi, *supra* note 54, at 887.
272. *Id.*; *see* COHEN, *supra* note 73, at 122 (this type of regime is supported by theories of efficiency).
273. *See* Lavi, *supra* note 54, at 931 (outlining different liability regimes for different types of social network platforms and arguing that adopting a "notice-and-takedown" safe haven regime provides a proper balance between constitutional rights and welfare considerations).

tion where the offending remarks appear. Due to the pace at which content is disseminated on social networks, this could prove to be an impossible mission.[274] Moreover, between the time the complaint is registered and the actual removal of content, the relevant content may continue to spread.

Applying "notice-and-takedown" or "right to be forgotten" regimes on search engines cannot confront the challenges of replication. Under these regimes, if the information was replicated to several websites and a search engine links to each and every one of them separately, the victim may have to contact each search engine about every single replication.[275] Moreover, removing a link to offensive content from search results may not stop the circulation of the content on various platforms.[276] This process disproportionately places the burden on victims without sufficiently mitigating harm.

Allowing complete immunity for intermediaries is even more under-inclusive. Under such a regime, the intermediary does not have a duty to remove defamation upon knowledge. Thus, even if the victim contacts the intermediary immediately after publication, the intermediary may still leave it on its platform. Consequently, the content may continue to spread and inflict harm.[277]

On the other hand, negligence liability is over-inclusive since negligence standards are generally open-ended. "Interpreting [negligence standards] involves cumbersome litigation and high administrative costs."[278] Difficulties experienced by courts in conducting cost-benefit analysis may result in inconsistency and uncertainty.[279] Negligence liability may also lead to "hindsight bias" and "outcome bias" because reasonable action is normally determined after the harm has already been inflicted.[280] "Consequently, courts

---

274. *See* RONSON, *supra* note 159, at 195–96.

275. The GDPR, *supra* note 248, art. 17.2, stipulates that "[w]here the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by controllers of any links to, or copy or replication of, those data." It is unclear what constitutes "reasonable steps" to erase replications. However, it is definitely clear that not erasing a replication after receiving a particular notice regarding it is unreasonable.

276. *See* Lavi, *supra* note 32, at 2654 (explaining that removing links from a search engine results only obscure the information and does not delete it altogether).

277. *See* Doug Lichtman & Eric Posner, *Holding Internet Service Providers Accountable*, 14 SUPREME CT. ECON. REV. 221, 222 (2006); SOLOVE, *supra* note 161, at 157–59.

278. Lavi, *supra* note 54, at 886.

279. *See* Guido Calabresi & Jon T. Hirschof, *Toward a Test for Strict Liability in Torts*, 81 YALE L.J. 1055, 1076 (1972).

280. Lavi, *supra* note 54, at 886; *see also* Yoed Halbersberg & Ehud Guttel, *Behavioral Economics and Tort Law*, *in* THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS & LAW 405, 411–12 (Eyal Zamir & Doron Teichman eds., 2014) ("Hindsight bias . . . distorts people's *ex post* assessments of the ex-ante probability and predictability of an event, given that this

may conclude that the [intermediary was] negligent even if he could not predict the harm *ex ante* and acted reasonably" at the time.[281] In the absence of a "safe haven," negligence liability may cast a heavy burden on the intermediary, resulting in a serious chilling effect on free speech.[282] Intermediaries can employ automatic algorithmic enforcement,[283] including Artificial Intelligence algorithms that are not sensitive enough to context,[284] to remove controversial content even before receiving a complaint[285] and still remain exposed to liability.[286] Moreover, algorithmic enforcement can impose costs on free speech; it can distort the public discourse by prioritizing certain types of content and erode democracy.[287] A negligence liability regime may also disincentivize innovations,

---

event has already happened . . . . The outcome bias is the tendency to perceive conduct that resulted in a bad outcome as more careless than the same conduct in cases where the bad outcome did not occur."); Baruch Fischhoff, *Hindsight Is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*, 1 J. EXPERIMENTAL PSYCH.: HUM. PERCEPTION & PERFORMANCE 288 (1975).

281. Lavi, *supra* note 54, at 886.

282. *See* KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, PRIVACY ON THE GROUND: DRIVING CORPORATE BEHAVIOR IN THE UNITED STATES AND EUROPE 242 (MIT Press ed., 2015) (explaining that vagueness in regulatory standards leads companies to implement higher standards of regulation); Danielle Keats Citron, *Extremist Speech and Compelled Conformity*, 93 NOTRE DAME L. REV. 1035–36 (2018) (in a related context of intermediary liability for incitement to terror, even planned legislation regarding intermediary liability caused the intermediary to overdo the removal of content by using digital tools in order to avoid potential liability).

283. *See, e.g.*, Federico Guerrini, *Facebook Will Flag and Filter Fake News In Germany*, FORBES (Jan. 16, 2017), https://www.forbes.com/sites/federicoguerrini/2017/01/16/facebook-will-flag-and-filter-fake-news-in-germany/ (describing a new technological screening tool that Facebook implemented due to the new German legislation that is expected to impose fines on the intermediary for fake news). For criticism on models of algorithmic enforcement, see Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016), which expounds on the growing use of algorithms by online intermediaries and the challenges of such cooperation enforcement. *See also* Benjamin Boroughf, *The Next Great YouTube: Improving Content ID to Foster Creativity, and Fair Compensation*, 25 ALB. L.J. SCI. & TECH. 95, 107 (2015).

284. *See* Natasha Duarte, Emma Llansó & Anna Loup, *Mixed Messages? The Limits of Automated Social Media Content Analysis* 1, 3 CTR. FOR DEMOCRACY & TECH. PAPER (2017), https://cdt.org/files/2017/12/FAT-conference-draft-2018.pdf ("Today's tools for analyzing social media text have limited ability to parse the meaning of human communication or detect the intent of the speaker."); *see also* Danielle Keats Citron & Neil M. Richards, *Four Principles for Digital Expression (You Won't Believe #3!)*, 95 WASH. U. L. REV. 1353, 1362 (2018).

285. Intermediaries may voluntarily apply best practices in order to be exempt from liability. *See* Niva Elkin-Koren & Orit Fischman-Afori, *Taking Users' Rights to the Next Level: A Pragmatist Approach,* 33 CARDOZO ARTS & ENT. L.J. 1, 36 (2015).

286. TUFECKI, TWITTER AND TEAR GAS, *supra* note 33, at 181 (explaining that disseminators of harmful speech find ways to bypass algorithmic enforcement for example by disseminating a print screen picture file that algorithms might find difficult to detect. Indeed, algorithms are improving, but disseminators of harmful expressions are likely to find ways to bypass algorithmic enforcement.).

287. For example, one of Facebook's strategies for combating fake news is using algorithms to prioritize content. On the use of algorithms for degrading and decreasing the visibility of fake news, see Daisuke Wakabayashi & Mike Issac, *In Race Against Fake News, Google and Facebook Stroll to the Starting Line*, N.Y. TIMES (Jan. 25, 2017), https://nyti.ms/35mUMoX.

such as mechanisms for sharing content, and thus reduce important advantages of digital markets, technological innovation, and the intermediary's freedom to conduct a business.[288] As a result, the positive externalities of sharing content are likely to decrease.

The above examination of overall immunity, "notice-and-takedown," and negligence liability regimes for regulating secondary intermediary liability reveals that they are either over- or under-inclusive. These regimes may, in fact, cause disproportionate chilling effects or allow extensive reputational harm and infringement of public interest. Moreover, common liability regimes alone are insufficient to meet the challenges posed by the prevalent dissemination of false rumors and defamation on social networks and decentralized private platforms.[289] While "notice-and-takedown" is preferable to other regimes for regulating intermediary liability on social network platforms, complementary mechanisms must also be incorporated in order to be effective. It should be noted that the harm of negative false rumors extends beyond the private individual's reputation and also falls into the public interest.[290] Therefore, policy makers should develop more tools to protect the public interest.

The following Section will describe the changing role of intermediaries in a brave new technological world. It will argue that the change in the intermediary's role requires a new concept of their accountability. As a first step, the Article will propose that the law should adapt the safe haven regime to the new technological reality of sharing. It proposes a new framework for content regulation. This framework is intended to mitigate the damage caused by dissemination of false rumors, defamation, and fake news, while preserving the benefits of dissemination and balancing the values at the base of intermediary liability.

---

288. The ambiguity regarding liability in Europe, reviewed in Section I.3.B above, has probably led many intermediaries to switch off readers' comments. *See* Paul McNally, *Guardian Digital Chief: Killing off Comments 'a Monumental Mistake'*, NEWS REWIRED (Mar. 2, 2015), https://www.newsrewired.com/2015/02/03/guardian-digital-chief-killing-off-comments-a-monumental-mistake. A negligence regime might also lead some intermediaries to avoid mechanisms that enable sharing at the click of a button. *See id.*

289. *See* MINA, *supra* note 99, at 126.

290. *See* Ben Shahar, *supra* note 11 (labeling this as "data pollution content" and proposing to use administrative and criminal law tools modeled on environmental law regulation). This Article will propose different solutions that focus on design.

D. *Reevaluating the Role and Obligations of Intermediaries in Light of Technological Developments*

People once thought of the internet as a sovereign-free medium controlled from the "bottom-up" by users without intermediation where anyone could publish anything without prior editing.[291] Instead, the internet simply shifted intermediation by creating new media gatekeepers.[292] These intermediaries are not mere conduits, as they control the flow of information.[293] While it seems as if everyone "can publish freely and instantly online," many intermediaries in fact "actively curate the content" that their users post on their platforms.[294] Intermediaries can organize the flow of information, promote or withhold ideas, and influence social dynamics.[295] "They act as centers for disseminating information and possess an essential role in directing the attention of users."[296] For example, intermediaries moderate user-generated content[297] by using various strategies that are hidden from public view,[298] with insufficient transparency.[299] Different intermediaries have different

---

291. Lavi, *supra* note 53, at 11–12 (citing Barlow, *supra* note 20). Barlow started the spirit of wide-eyed techno-utopianism.

292. *See* David S. Evans, *Governing Bad Behavior by Users of Multi-Sided Platforms*, 27 BERKELEY TECH. L.J. 1201, 1201 (2012) (explaining that platforms often develop government mechanisms through which they monitor and manage "bad" behavior, thereby acting as gatekeepers); MARANTZ, *supra* note 39, at 70. *See generally* COHEN, *supra* note 73, at 37–38, 75 (explaining that some aspects the conception of "technologies of freedom" changed beyond recognition and today's networked digital information infrastructure have different and more complicated affordances).

293. Olivier Sylvain, *Intermediary Design Duties*, 50 CONN. L. REV. 203, 220 (2018) (explaining that because YouTube structures, sorts and sometimes sells users' data, it is not a passive conduit); *see also* Balkin, *supra* note 183, at 2297–98 (2014); Derek E. Bambauer, *Middlemen*, 64 FLA. L. REV. 64, 65 (2013).

294. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1599 (2018); *see also* SARAH T ROBERTS, BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA 93–161 (2019) (describing how different types of low-wage human contractors moderate content); Rory Van Loo, *Federal Rules of Platform Procedure*, 87 U. CHI. L. REV. (forthcoming 2020) (manuscript at 10) (on file with the author) (describing how Facebook cut off vital avenues for speech and sharing information, account termination and even deprives a user of valuable property without adequate transparency).

295. Michal Lavi, *Online Intermediaries: With Power Comes Responsibility*, JOLT DIG. (May 11, 2018), https://jolt.law.harvard.edu/digest/online-intermediaries-with-power-comes-responsibility.

296. *Id.*

297. *See* TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 5–6 (2018); Kate Klonick, *The Facebook Oversight Board Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L.J. 2418, 2427 (2020) ("Content moderation is the industry term for a platform's review of user-generated content posted on its site and the corresponding decision to keep it up or take it down.").

298. Klonick, *supra* note 297, at 2427.

299. *See id.* at 2418 (describing how Facebook built institutions for oversight and explaining that only after facing outside pressure from media, government, and the civil society did

attitudes towards moderation and different rules of community.[300] Intermediaries can structure user participation on their platforms and push users to disclose and share information.[301] Twitter for example, uses their algorithms to "influence what is viewed, what is valued, and what is disseminated and re-disseminated by users."[302] They collect information on users, personalize content,[303] and influence users' behavior, decision-making processes, social dynamics, the content they create,[304] and even how they participate in democracy.[305] For that reason, intermediaries have been dubbed the "New Governors of online speech."[306]

As technologies advance and the role of the intermediary as moderator of the flow of information becomes a fundamental aspect of any platform, the duties of intermediaries should be reconsidered.[307] Reevaluating intermediaries' roles and duties is of par-

---

Facebook insert more transparency into its moderation practices and dedicate more resources to enforcement). These measures however are insufficient. Moreover, other media giants might be even less transparent regarding their moderation practices.

300. *See* Shannon Bond, *Critics Slam Facebook but Zuckerberg Resists Blocking Trump's Posts*, NPR (June 11, 2020), https://n.pr/37mIoqm ("When Trump tweeted an identical message, Twitter took the novel step of hiding the tweet behind a warning label, saying it broke its rules against glorifying violence. Zuckerberg saw it differently. Even though he was personally disgusted by the president's inflammatory rhetoric, he said, the post did not break Facebook's rules against inciting violence."); Douek, *supra* note 182, at 5 (explaining that major platforms, are crafted around two different precepts: proportionality and probability).

301. *See supra* Section I.D.

302. Lavi, *supra* note 295; *see also* Alex Hern, *Twitter Hides Donald Trump Tweet for 'Glorifying Violence*,' THE GUARDIAN (May 29, 2020), https://www.theguardian.com/technology/2020/may/29/twitter-hides-donald-trump-tweet-glorifying-violence.

303. *See* ZUBOFF, *supra* note 115, at 466 (describing the rise of surveillance capitalism); FRISCHMANN & SELINGER, *supra* note 130, at 117–18 (describing the emotional cognition experiment that shows that intermediaries can control what is seen and what is disseminated). In another related context, Facebook allowed advertisers to target advertisements on specific topics to hate groups. *See* Julia Angwin, Madeleine Varner & Ariana Tobin, *Facebook Enabled Advertisers to Reach 'Jew Haters'*, PROPUBLICA (Sept. 14, 2017), https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters; Kerri A. Thompson, *Commercial Clicks: Advertising Algorithms as Commercial Speech*, 21 VAND. J. ENT. & TECH. L. 1019 (2019).

304. *See* GILLESPIE, *supra* note 297, at 23 ("Platforms may not shape the public discourse by themselves, but they do shape the shape of the public discourse. And they know it.").

305. *See* Jonathan Zittrain, *supra* note 55, at 335; VAIDHYANATHAN, *supra* note 131, at 87–89 (describing how personalized content affected the 2016 election); Carole Cadwalladr & Emma Graham-Harrison, *Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach*, THE GUARDIAN (Mar. 17, 2018), https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

306. Klonick, *supra* note 294, at 1603.

307. *See* Thomas E. Kadri, *Digital Gatekeepers*, 99 TEX. L REV. (forthcoming 2021) (manuscript at 2–3) ("[F]ollowing years of laissez-faire attitudes in legislatures, lawmakers are looking for ways to regulate the technology companies that exert so much influence over our lives."). Recently, even Facebook founder Mark Zuckerberg has conceded that the internet needs new rules. Mark Zuckerberg, Opinion, *The Internet Needs New Rules. Let's Start in These Four Areas,* WASH. POST (Mar. 30, 2019), https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html.

ticular importance, especially in light of the recent attack on § 230.[308]

As I have proposed elsewhere, intermediaries that influence users to publish and share false rumors and defamation, in particular, can be held liable under a contributory liability regime.[309] Intermediaries that publish defamation themselves, or mix their own content with users' defamatory content, might bear direct liability for defamation.[310] But what about the liability of intermediaries that make no particular effort to promote or publish harmful content? Should the law impose an obligation on intermediaries just because they sway influence on users to publish more content, even if they have not promoted or repeated offensive content in particular?

The power of intermediaries to shape the flow of information has inspired debate in legal scholarship. "Recent scholarship acknowledges that twenty-first century intermediaries . . . cannot be treated as mere passive conduits and that their role and duties should be reconceptualized."[311] Thus, new concepts of the intermediary's role are being developed. Different scholars have observed the influences of intermediaries in different ways and have proposed different types of legal obligations.

Even though intermediaries are private entities, some scholars have proposed that since they control the information infrastructures that serve the public, they should be treated as public forums, or at least hybrid bodies.[312] These scholars argue that intermediaries should be treated as state actors and therefore subjected to the First Amendment and other basic public law standards.[313] This per-

---

308. See Exec. Order, *supra* note 196, and the new bills, *supra* note 208, that propose to amend § 230.

309. *See* Lavi, *supra* note 53; Lavi, *supra* note 141, at 478.

310. Lavi, *supra* note 25.

311. Lavi, *supra* note 141, at 544; *see also, e.g.*, Kyle Langvardt, *Regulating Online Content Moderation*, 106 GEO. L.J. 1353, 1373 (2018).

312. Kyle Langvardt, *A New Deal for the Online Public Sphere*, 26 GEO. MASON L. REV. 341, 380–81 (2018) (proposing that nonstate regulators such as online platforms can be perceived as state agencies); K. Sabeel Rahman, *Private Power, Public Values: Regulating Social Infrastructure in a Changing Economy*, 39 CARDOZO L. REV. 1621, 1668 (2018) (proposing to apply public utilities concept on online platforms); Orit Fischman-Afori, *Online Rulers as Hybrid Bodies: The Case of Infringing Content Monitoring*, 23 U. PA. J. CONST. L. (forthcoming 2020) (proposing that online platforms should be treated as hybrid bodies and subject them to public law standards).

313. Rahman, *supra* note 312, at 1687. It should be noted that profiles of government entities and government representatives are already treated as public forums. *See, e.g.*, Davison v. Randall, 912 F.3d 666 (4th Cir. 2019). For example, the court ruled that then-President Donald Trump could not block Twitter followers due to their dissenting views because to do so is a violation of their First Amendment right to participate in a "designated public forum." *See* Knight First Amend. Inst. at Columbia Univ. v. Trump, 302 F. Supp. 3d 541 (S.D.N.Y. 2018), *aff'd*, 928 F.3d 226 (2d Cir. 2019).

ception of intermediaries was rejected by the Ninth Circuit decision in *Prager University v. Google LLC*; the court concluded that YouTube is not a public forum.[314] Such a perception, however, is somewhat reflected in the recent executive order that declared: "It is the policy of the United States that large online platforms, such as Twitter and Facebook, as the critical means of promoting the free flow of speech and ideas today, should not restrict protected speech."[315]

As Professor Balkin has explained, ultimately, imposing the full spectrum of public forum obligations on intermediaries is undesirable and would actually make things worse.[316]

> It would do nothing to prevent third parties from using social media to manipulate end users, stoke hatred, fear, and prejudice, or spread fake news. And because social media would be required to serve as [a] neutral public forum [and obligated to equality], they could do little to stop this.[317]

Even if social media platforms desist from curating feeds, they would still collect and harvest data on their end-users, either directly or by using third parties, such as mobile apps, as recently leaked documents from Facebook demonstrate.[318] This data, in

---

314. 951 F.3d 991, 995 (9th Cir. 2020). Prager University claimed that, by classifying some of their videos as "Restricted Content," YouTube attempted to silence "conservative viewpoints and perspectives on public issues," censored their content, and violated their First Amendment rights. The Ninth Circuit upheld the dismissal of the case concluding that: "Despite YouTube's ubiquity and its role as a public-facing forum, it remains a private forum, not a public forum subject to judicial scrutiny under the First Amendment." *Id.*

315. Exec. Order, *supra* note 196, at 34,081. A prior version of this sentence referenced the public forum doctrine. *See* Goldman, *supra* note 205.

316. Jack M. Balkin, *Fixing Social Media's Grand Bargain* (Hoover Working Group on Nat'l Sec., Tech. & L., Aegis Series Paper No. 1814) ("Treating social media companies as public forums or public utilities is not the proper cure. It may actually make things worse.") [hereinafter Balkin, *Grand Bargain*]; Balkin, *supra* note 170 ("Converting all large social media companies into public utilities does not solve the problems I mentioned above, because it does not provide diverse affordances, value systems, and innovations.").

317. Balkin, *Grand Bargain, supra* note 316, at 6; *see also* Langvardt, *supra* note 311, at 1367 ("[T]he more significant difficulty with applying the state action doctrine to the platforms lies in the fact that internet platforms can "evict" unwanted speakers without involving the courts."); Citron & Franks, *supra* note 194, at 62–63.

318. Balkin, *Grand Bargain*, *supra* note 316, at 6; Sebastian Klovig Skelton & Bill Goodwin, *Lawmakers Study Leaked Facebook Documents Made Public Today*, COMPUT. WKLY. (Nov. 6, 2019), https://www.computerweekly.com/news/252473540/Lawmakers-study-leaked-Facebook-documents-made-public-today; *Facebook Sold a Rival-Squashing Move as Privacy Policy, Documents Reveal*, THE GUARDIAN (Nov. 6, 2019) https://www.theguardian.com/us-news/2019/nov/06/facebook-privacy-switcharoo-plan-emails.

turn, could be sold to third parties, who could use it on their sites or elsewhere and influence the flow of information.[319]

A second proposal is to view intermediaries as a hybrid between a conduit and a media company.[320] Intermediaries not only host content but they also connect users, organize content, make content searchable, and recommend relevant content to users through algorithms. Their algorithms select certain content and gives it preference over other content based on judgments of relevance and considerations of keeping users on the site.[321] Intermediaries create an ecosystem of networked journalism through personalized recommendations and contribute to how news is made.[322] They are a key pathway to news, even surpassing print newspapers as a news source.[323] Arguably, as the similarities between intermediaries and media companies grow, intermediaries should be subjected to some of the professional norms and standards that apply to traditional media.[324] Indeed, some intermediaries already apply professional standards and restrict specific types of content on their platforms in their terms of service and community standards.[325] Yet, the law still has a role in shaping the framework.[326]

A third proposal is the concept of information fiduciaries. This approach likens the obligation of intermediaries towards user information to the fiduciary duties of doctors and lawyers towards patients and clients. "Just as the law imposes special duties of care,

---

319. Balkin, *Grand Bargain*, *supra* note 316, at 6 ("[T]reating social media as public forums would only affect the ability of social media themselves to manipulate end users. It would do nothing to prevent third parties from using social media to manipulate end users, stoke hatred, fear, and prejudice, or spread fake news.").

320. *See generally* Mary Louise Kelly, *Media or Tech Company? Facebook's Profile Is Blurry*, NPR (Apr. 11, 2018), https://www.npr.org/2018/04/11/601560213/media-or-tech-company-facebooks-profile-is-blurry (explaining that lawmakers and regulators have a hard time determining whether Facebook is a media or tech company).

321. *See, e.g.*, GILLESPIE, *supra* note 297, at 43 ("As soon as Facebook changed from delivering a reverse chronological list of materials that users posted on their walls to curating an algorithmically selected subset of those posts in order to generate a News Feed, it moved from delivering information to producing a media commodity out of it.").

322. Erin C. Carrol, *Platforms and the Fall of the Fourth Estate: Looking Beyond the First Amendment to Protect Watchdog Journalism*, 79 MD. L. REV. 529, 531–32 (2020).

323. Katherine Schaeffer, *U.S. Has Changed in Key Ways in The Past Decade, from Tech Use to Demographics*, PEW RSCH. CTR. (Dec. 20, 2019), https://www.pewresearch.org/fact-tank/2019/12/20/key-ways-us-changed-in-past-decade/ ("Social media is now a key pathway to news for Americans. In 2018, for the first time, social media sites surpassed print newspapers as a news source for Americans.").

324. *See* GILLESPIE, *supra* note 297, at 43; Balkin, *Grand Bargain*, *supra* note 316, at 8 (giving examples of professional standards that social media should live up to, such as adhering to professional standards of journalistic ethics).

325. *See, e.g.*, *Community Standards: Bullying and Harassment*, FACEBOOK, https://www.facebook.com/communitystandards/bullying (last visited May 11, 2020) ("[Facebook will] remove content that's meant to degrade or shame, including, for example, claims about someone's sexual activity.").

326. *See* Balkin, *Grand Bargain*, *supra* note 316.

confidentiality, and loyalty on doctors [and] lawyers [with regard to] their patients and clients, . . . it [should] impose special duties on [intermediaries] such as Facebook, Google, and Twitter [towards] their users."[327] Intermediaries resemble fiduciaries because, much like lawyers and doctors, they receive—and even actively collect—personal information and are trusted to treat it with care. Intermediaries obtain information that their users knowingly disseminate on their platforms and actively collect incidental information on their users' engagement on the platform that leaves digital traces.[328] Therefore, some have argued that the law should impose duties of care, confidentiality, and loyalty on intermediaries and limit how they can profit from their users and beneficiaries.[329] The nature of fiduciary obligations should depend on the nature of the relationship and the potential risk for abuse in using the information by the more powerful party to the relationship.[330] In our context, "[i]ntermediaries should neither breach user trust nor take actions that users would reasonably consider unexpected or abusive."[331] As information fiduciaries, intermediaries should have a duty not to utilize user data to influence or even manipulate them.[332] This view strives to impose on intermediaries duties to operate their platforms with good faith, respect for users, and non-manipulation.[333] "It should be noted that the information fiduciary [approach] raises challenges regarding its feasibility, enforceability and scope."[334]

---

327. Lina M. Khan & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497, 498 (2019); *see also* Balkin, *Grand Bargain*, *supra* note 316, at 12.

328. *See* Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World,* 4 GEO. L. TECH. REV. 1, 25–26 (2019) ("[B]oth the information we knowingly disseminate about ourselves when we visit websites, make online purchases, and post photographs and videos on social media, and the information we unwittingly provide (e.g., when those websites record data about how long we spend reading them, where we are when we access them, and which advertisements we click on) reveals a great deal about who we are, what interests us, and what we find amusing, tempting, and off-putting."); TUROW, *supra* note 116, at 34 (explaining that intermediaries can collect data on consumers online by tracking browsing activities, clicks, cookies and actual purchases); ZUBOFF, supra note 115, at 80 ("[T]hese include websites visited, psychographics, browsing activity and information about previous advertisements that the users have been shown, selected and/or made purchases after viewing.").

329. *E.g.*, Balkin, *Grand Bargain*, *supra* note 316, at 12.

330. Jack M. Balkin, *The Fiduciary Model of Privacy*, 134 HARV. L. REV. F. 11,15 (2020).

331. Lavi, *supra* note 53, at 68; *see also* Balkin, *supra* note 330, at 13; Jack M. Balkin, *The First Amendment in the Second Gilded Age*, 66 BUFF. L. REV. 979, 1006–08 (2018); Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1229 (2016).

332. Lavi, *supra* note 53, at 68 (citing Jack M. Balkin, *Free Speech Is a Triangle,* COLUM. L. REV. 2011, 2049 (2018)); Balkin, *Grand Bargain, supra* note 316, at 14.

333. Lavi, *supra* note 141, at 545 n.458 (citing Jack M. Balkin & Jonathan Zittrain, *A Grand Bargain to Make Tech Companies Trustworthy*, THE ATLANTIC (Oct. 3, 2016), https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/).

334. Lavi, *supra* note 32, at 2641 n.186. For a critique of the theory of information fiduciaries, see Khan & Pozen, *supra* note 327.

Unlike the EU GDPR data protection obligations,[335] the concept of information fiduciaries does not rely on structuring privacy by obtaining user consent for individual transactions.[336] "Rather, the fiduciary approach holds digital fiduciaries to obligations of good faith and non-manipulation regardless of" their particular privacy policies.[337] Defining the appropriate concept for the intermediary is beyond the scope of this Article, which focuses on the dissemination of false rumors. Be that as it may, the growing influence of intermediaries on the flow of information may justify the imposition of obligations on intermediaries that go beyond traditional liability regimes. Designing the appropriate legal governance that should apply to online intermediaries is one of the most urgent legal challenges at this time. Many scholars believe it is high time to change the overall immunity regime applied to intermediaries and adapt it to the high degree of influence they exert over users.[338] In the context of intermediary accountability for harmful false rumors, this Article proposes imposing concrete obligations upon intermediaries to mitigate the harm of dissemination of false rumors. In contrast to the executive order of Donald Trump,[339] the 45th President of the United States, the proposal does not undermine intermediaries' practices of moderation and preserves freedom of expression.

III. MEETING THE CHALLENGES OF SPREADING FALSEHOODS ON SOCIAL NETWORKS

Intermediaries use different strategies to facilitate dissemination of content in their attempts to increase profits.[340] As previously explained, content dissemination provides multiple benefits.[341] The challenge becomes how to allow the free flow of information while simultaneously preventing the dissemination of harmful content. To meet this challenge, I will focus on the design stage of a plat-

---

335. GDPR, supra note 248. *See generally infra* Part I.

336. Balkin, *Grand Bargain*, *supra* note 316, at 14 ("[C]ontractual models will prove insufficient if end users are unable to assess the cumulative risk of granting permission and therefore must depend on the good will of data processors. The fiduciary approach to obligation does not turn on consent to particular transactions. . . .").

337. *Id.*

338. *See, e.g.*, Sylvain, *supra* note 293, at 258 (2018); Olivier Sylvain, *Discriminatory Designs on User Data*, KNIGHT FIRST AMENDMENT INST. COLUM. U. (2018), https://knightcolumbia.org/content/discriminatory-designs-user-data ("[T]hese developments undermine any notion that online intermediaries deserve immunity because they are mere conduits for, or passive publishers of, their users' expression.").

339. *See* Exec. Order, *supra* note 196.

340. *See id.*

341. *See id.*

form's lifecycle. One solution focuses on reducing irresponsible dissemination *ex ante* at the stage of the user's decision to share content. The second solution strives to mitigate harm *ex post facto*.

A. *Protecting the Right to Reputation and Public Interest by Design*

In the book *Code Version 2.0*, Lawrence Lessig identified four key forces that regulate the online environment.[342] First, laws regulate and constrain activities and can impose sanctions when activity violates them. For example,"[c]opyright law, defamation law, and obscenity laws all continue to threaten ex post sanction for the violation of legal rights" online.[343] Second, norms restrict activities by stigmatizing violations. For example, "talk about Democratic politics in the alt.knitting newsgroup, and you open yourself to flaming."[344] In other words, violation of the norms in a newsgroup increases the likelihood of encountering insults and hostile aggressive interactions.[345] Third, the market limits activities by price-setting, thus high prices can constrain access of individuals to goods and services.[346] Fourth, technologies can constrain by "code," meaning that the software and hardware design can define freedoms online, affect user choices, and regulate their interactions. For example, some websites require a person to enter a password and identify himself before gaining access; on other sites, a person can enter whether identified or not.[347] These four governance systems all interact simultaneously.[348]

In recent years, there has been an increasing use of technology-based solutions to prevent harm inflicted by the free flow of information. Architecture of online platforms, namely the way they are designed, can involve the creation of structures to prevent harm from arising and shape attitudes towards violations of law and

---

342. *See* LAWRENCE LESSIG, CODE: VERSION 2.0 121–125 (2006).
343. *Id.* at 124.
344. *See, e.g.*, *id.* at 124–26.
345. *See id.* at 124 ("[T]alk too much in a discussion list, and you are likely to be placed on a common bozo filter. In each case, a set of understandings constrain behavior, again through the threat of ex post sanctions imposed by a community."). For expansion on flaming, see generally Patrick B. O' Sullivan & Andrew J. Flanagin, *Reconceptualizing 'Flaming' and other Problematic Messages,* 5 NEW MEDIA & SOC'Y 69 (2003).
346. *Id.*
347. *See id.* at 125; Tal Zarsky, *Social Justice, Social Norms and the Governance of Social Media*, 35 PACE L. REV. 138, 139 (2015) (adjusting the model to social media).
348. *See* COURTNEY BOWMAN, ARI GESHER, JOHN K. GRANT, DANIEL SLATE & ELISSA LERNER, THE ARCHITECTURE OF PRIVACY: ON ENGINEERING TECHNOLOGIES THAT CAN DELIVER TRUSTWORTHY SAFEGUARDS 13 (2015) (referring to the interaction of code and law as "East Coast" code and "West Coast" code).

norms *ex ante*,[349] as well as promote values.[350] Studies have emphasized the power of architecture to account for human values and technology user rights "in a principled and comprehensive manner throughout the design process."[351] Other studies have focused on strategies of influence, starting with behavioral influences of the design on decision-making and continuing with the role of code in outlining possibilities by altering the platform's design to "make certain conduct more difficult or costly."[352]

Decisions made by engineers can unleash new technology not previously foreseen by the legislator, which may affect fundamental rights.[353] Scholarly work has already explored the influence of technological governance systems and their potential to protect privacy. This concept of privacy by design was developed into a philosophy that focuses on regulation of the technological design *ex ante* instead of providing *ex post* remedies to victims of dissemination of harmful information.[354] Researchers have described how to make privacy-protective features a core part of functionality and accommodate threats to privacy.[355] Scholars have also noted that the primary challenges of privacy by design are enhancing the specification and incentivizing firms to adopt this approach.[356]

---

349. *See* DANIEL SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 100 (2004).

350. Mulligan & Bamberger, *supra* note 110, at 701, 708–09. ("Designing technology to "bake in" values offers a seductively elegant and effective means of control.").

351. Deirdre K Mulligan & Jenifer King, *Bridging the Gap Between Privacy and Design*, 14 U. PA. J. CONST. L. 989, 1019 (2012) (quoting Batya Friedman, Peter H. Kahn, Jr., & Alan Borning, DEP'T OF COMPUT. SCI. & ENG'G, UNIV. OF WASH., CSE TECHNICAL REPORT NO. 02-12-01, VALUE SENSITIVE DESIGN: THEORY AND METHODS (2002)). For instance, " 'Value-Sensitive Design' is an approach that advocates identifying human needs and values and taking them into account in the design process." Lavi, *supra* note 141, at 553 n.504 (citing Noemi Manders-Huits & Jeroen van den Hove, *The Need for Value-Sensitive Design of Communication Infrastructures*, *in* EVALUATING NEW TECHNOLOGIES 51, 54–55 (Paul Sollie & Marcus Düwell eds., 2009); Mulligan & King, *supra*).

352. Ryan Calo, *Code, Nudge, or Notice?*, 99 IOWA L. REV. 773, 775, 778 (2014) (noting that these strategies are blended and recombined); *see also* TUFECKI, TWITTER AND TEAR GAS, *supra* note 33, at 264–65 (addressing the power of technology to target content and facilitate viral spread).

353. Lavi, *supra* note 141, at 553.

354. *See* CHRIS JAY HOOFNAGLE, FEDERAL TRADE COMMISSION PRIVACY LAW & POLICY 190–91 (2016); *see also, e.g.*, KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, PRIVACY ON THE GROUND: DRIVING CORPORATE BEHAVIOR IN THE UNITED STATES AND EUROPE 32, 178 (2015); BOWMAN ET AL., *supra* note 348; ANN CAVOUKIAN, PRIVACY BY DESIGN 1, 2 (2009).

355. *See, e.g.*, BOWMAN ET AL., *supra* note 348; Ira S. Rubinstein, *Regulating Privacy by Design*, 26 BERKELEY TECH. L.J. 1409, 1420–21 (2011); Frederic Stutzman & Woodrow Hartzog, *Obscurity by Design*, 88 WASH. L. REV. 385, 395, 402–417 (2013) (referring to a narrow approach to privacy (obscurity of data) and overview strategies to protect privacy by design); Serge Egelman, Janice Tsai, Lorrie Cranor & Alessandro Acquisti, *Timing Is Everything? The Effects of Timing and Placement of Online Privacy Indicators*, PROC. A.C.M. S.I.G.C.H.I. CONF. ON HUM. FACTORS IN COMPUT. SYS. 319 (2009).

356. *See* Rubinstein, *supra* note 355, at 1416; Stutzman & Hartzog, *supra* note 355, at 392.

Regulators around the world have discovered the benefits of privacy by design. They have set forth guidelines and promoted legal regulation that includes privacy by design, alongside efforts to incentivize stakeholders to adopt this approach as part of their business models.[357] A central example is Article 25 of the GDPR that addresses "data protection by design and default,"[358] building privacy-friendly systems starting at the beginning of the design process.[359] Accordingly, at the stage of the system development, and at the time of processing, controllers must "implement appropriate technical and organizational measures" in order to "protect the rights of data subjects."[360] "Data protection by default" is required to assure that data that is unnecessary for processing is not gathered.[361] Examples of data protection by design are "anonymisation and pseudonymisation of personal data, a data minimisation approach during processing and storing data, storage limitation, transparency regarding processing and limited access to personal data."[362]

"The GDPR protects data of EU citizens, but it [also] applies to non-EU companies that offer goods or services to EU consumers."[363] Thus, the GDPR can also affect data protection in the United States and throughout the world. Furthermore, the GDPR contains a "threshold test for international transfers of personal data to [non-member states] and a legal basis for blocking data exports to [states] that do not meet this standard."[364] The threshold for extraterritorial transmissions is the "*adequacy*" of data protection in the foreign jurisdiction.[365] With regard to transmissions to the United States, instead of an adequacy determination, the Europe-

---

357. *See, e.g.*, FTC, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS (2012), http://ftc.gov/os/2012/03/120326privacyreport.pdf; HOOFNAGLE, *supra* note 354, at 191 ("[T]he FTC is embracing privacy by design."); *see also* A Comprehensive Approach On Personal Data Protection in the European Union, EUR. PARL. DOC. (COM 609) 12 (2010); Edwards & Veale, *supra* note 248, at 77.

358. GDPR, supra note 248, at art. 25.

359. *See* Edwards & Veale, *supra* note 248, at 77 (explaining that by doing so, it recognizes that "a regulator cannot do everything by top-down control, but . . . controllers must themselves be involved in the design" of systems that minimize invasion of privacy).

360. GDPR, *supra* note 248, at art. 25.

361. Edwards & Veale, *supra* note 248, at 77 (quoting GDPR, *supra* note 248, art. 25).

362. Oliver Vettermann, *Self-Made Data Protection—Is It Enough? Prevention and After-care of Identity Theft*, 10 Eur. J.L. & Tech. § 4.2 (2019).

363. Lavi, *supra* note 141, at 563 n.559.

364. Paul M. Schwartz, *Global Data Privacy: The EU Way*, 94 N.Y.U. L. REV. 771, 784 (2019).

365. *Id.* ("In Article 45, the GDPR requires that the Commission consider a long list of factors in assessing the adequacy of protection, including 'the rule of law, respect for human rights and fundamental freedoms, relevant legislation, both general and sector, . . . as well as the implementation of such legislation, data protection rules, professional rules and security measures.' " (quoting GDPR, *supra* note 248, at art. 45(2)(a)).

an Union and the United States have reached an arrangement called "Privacy Shield," a voluntary private sector compliance program.[366] This bilateral agreement "present[s] a list of substantive EU principles for American companies to follow voluntarily."[367] Recently, however, the ECJ in Luxembourg struck down the Privacy Shield in the case of *Data Protection Commissioner v. Facebook Ireland Ltd.*,[368] determining that the Privacy Shield agreement did not limit the U.S. authorities' access to data "in a way that satisfies requirements that are essentially equivalent to those required . . . under EU law." The impact of the ruling is not yet clear. The GDPR, thus, has global impact today, more than ever, and the principles of privacy by design can influence the engineering of privacy outside of Europe.[369]

As mentioned, most discussions on behavioral influences of design and the technological constraints of code have focused on privacy protection. This Article proposes to adopt this strategy for preventing impulsive dissemination of false rumors, defamation, and fake news, and suggests incentives that would lead intermediaries to adopt these proposals. Platform design and code have promising potential to protect personal reputation from harm. In fact, the same technologies, strategies, and principles used by intermediaries to promote dissemination of content can be utilized to mitigate harm inflicted by the dissemination of falsehoods. Due to the potential of these strategies for the protection of reputations and the public interest in general, engineers, managers, and policy makers should develop a concept of "Reputation and Public Interest-by-Design." The importance of design for the protection of reputations is reinforced by the insufficiency of current law to accommodate the challenge of fast-spreading falsehoods.[370] This concept not only protects the private interests of victims of falsehoods, but it also has a role in promoting the public interest in truthful information.

---

366. It should be noted that in *Schrems v. Data Protection Commissioner*, the ECJ declared that this safe harbor was invalid. Case C-362/14, ECLI:EU:C:2015 ¶ 98 (Oct. 6, 2015). Following the decision in that case, the United States and the European Union came to an agreement on the Privacy Shield. *See* Commission Implementing Decision of July 12, 2016 Pursuant to Directive 95/46/EC of the European Parliament and of the Council on the Adequacy of the Protection by the EU–U.S. Privacy Shield, 2016 O.J. (L 207) 1, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:207:TOC.

367. *See* Schwartz, *supra* note 364, at 802.

368. Case C-311/18, ECLI:EU:C:2020:559 ¶ 185 (July 16, 2020).

369. Beata A. Safari, *Intangible Privacy Rights: How Europe's GDPR Will Set a New Global Standard For Personal Data Protection*, 47 SETON HALL L. REV. 809, 816–20 (2017); Schwartz, *supra* note 364, at 777 ("[P]rinciples found in the GDPR, such as data portability and the 'right to be forgotten,' are already influencing laws outside Europe."); *see also, e.g.*, Rustad & Koenig, *supra* note 251, at 420.

370. *See supra* Part II.

B. *From Nudges to Accountable Dissemination of Information*

In their seminal book *Nudge*, Richard Thaler and Cass Sunstein show that policy makers can arrange decision-making contexts and thus nudge individuals to change their behavior.[371] A nudge is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives."[372] Under this idea of 'libertarian paternalism,'[373] the individual who designs the environment for people's decision making—the "choice architect"[374]—may anticipate their behavior, respond to this prediction, and direct them to act in a certain way.[375] This concept is also applicable to preventing third-party harm; for example, nudges can be used to reduce "texting and driving" and, therefore, reduce accidents that would cause harm to third parties.[376] "Advocates of the nudge approach believe that choice-preserving alternatives are preferable to mandates."[377] The nudge concept has attracted controversy, objections and ethical concerns; yet, it has also achieved widespread recognition among policy makers and has even led to reforms.[378]

Sunstein and Thaler reviewed many examples of nudges.[379] For example, a charity debit card that keeps a record of an individual's donations and ensures that their bank adds their donations to their end-of-year IRS statement makes donating more attractive and incentivizes individuals to donate more.[380] A nudge commit-

---

371. Lavi, *supra* note 53, at 4 (citing THALER & SUNSTEIN, *supra* note 30); *see also* CASS R. SUNSTEIN, WHY NUDGE? THE POLITICS OF LIBERTARIAN PATERNALISM (2014) (delving deeper into the debate about the rationales and objections of nudges).

372. THALER & SUNSTEIN, *supra* note 30, at 6.

373. SUNSTEIN, *supra* note 371, at 358–59 (emphasizing the idea of continuum and recognizing that approaches that impose high psychological costs are not as soft as approaches that impose low material costs).

374. *See* Richard H. Thaler, Cass R. Sunstein & John P. Balz, *Choice Architecture*, *in* THE BEHAVIORAL FOUNDATIONS OF PUBLIC POLICY 428 (Eldar Shafir ed., 2012).

375. Lavi, supra note 53, at 4.

376. *See* SUNSTEIN, *supra* note 371 at 9, 44; Christopher McCrudden & Jeff King, *The Dark Side of Nudging: The Ethics, Political Economy, and Law of Libertarian Paternalism*, *in* CHOICE ARCHITECTURE IN DEMOCRACIES, EXPLORING THE LEGITIMACY OF NUDGING 67, 93 (Alexandra Kemmerer, Christoph Möllers, Maximilian Steinbeis & Gerhard Wagner eds., 2016) (arguing that by referring to "texting while driving" and "fuel standards" as areas where nudging is appropriate, harm can be prevented).

377. Lavi, *supra* note 53, at 8; *see also* Calo, *supra* note 352, at 783; Cass R. Sunstein, *Nudges vs. Shoves*, 127 HARV. L. REV. F. 210 (2014).

378. Lavi, *supra* note 53, at 9; *see also* Thaler, s*upra* note 373, at 331; Cass R. Sunstein, *Do People Like Nudges?*, 68 ADMIN. L. REV. 177 (2016).

379. *See* THALER & SUNSTEIN, *supra* note 30, at 237.

380. *See* Yang Wang, Pedro Giovanni Leon, Xiaoxuan Chen, Saranga Komanduri, Gregory Norcie, Kevin Scott, Alessandro Acquisti, Lorrie Faith Cranor & Norman Sadeh, *From Facebook Regrets to Facebook Privacy Nudges*, 74 OHIO STATE L.J. 1307, 1319–23, 1331 (2013) (arguing that designing mechanisms that nudge users to consider the content and context of their online disclosures are efficient in helping individuals avoid regrettable online disclo-

ting people to a specific action, such as exercising more frequently, can help them achieve their goals.[381] By emailing announcements about failures or successes to family and friends, or monitoring a person's goals via a group blog, peer pressure "nudges" people to fulfil their goals.[382] A red warning light on a water filter or an air conditioner "nudges" individuals to replace the filter.[383] Devices and apps such as GPS and Waze can "nudge" users by showing them the most efficient route of driving.[384] Organizing the default settings on websites and the order of choices in online menus can also serve as a type of nudge.[385]

A central example of a nudge that is relevant to our context is a "civility check" on a message that a person intends to send. Each and every hour, people send out angry emails that they soon regret.[386] A "civility check" nudges individuals to file the message and wait a day before deciding whether to send it;[387] meanwhile, the individual might calm down and reconsider. Technology can facilitate this civility check by detecting whether an email is angry, cautioning him that it appears to be uncivil, and asking whether he really wants to send it. A stronger version might delay dissemination of uncivil emails by default and force individuals to invest extra efforts in order to bypass the delay, such as requiring the entering a social security number.[388] This nudge would lead individuals to reflect upon whether they truly wish to send the email.

Nudge-based strategies can be smoothly transplanted to social media because technology makes it easy to arrange decision-making contexts, compared to a brick-and-mortar environment,[389] and such strategies are becoming more powerful in platform-

---

sures and, thus, enhance their privacy); Leslie K. John, Alessandro Acquisti & George Loewenstein, *Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information*, 37 J. CONSUMER RSCH. 858 (2011).

381. Thus, whenever a user tweets, the intermediary informs him that his tweets will appear in the timelines of his followers. However, this message appears only after dissemination and the "status-quo-bias" reduces the likelihood of *ex post facto* deletion by the user. *See* THALER & SUNSTEIN, *supra* note 30, at 34 (explaining that people have a general tendency to stick with the current situation, this phenomenon is dubbed the "status-quo-bias").

382. *Id.*

383. *Id.* at 237.

384. *Id.* at 242.

385. *Id.* at 8.

386. *Id.* at 237.

387. *Id.* at 237.

388. *Id.* at 237–38.

389. Programming platforms is easier than designing brick-and-mortar architecture. *See* Yochai Benkler, *Degrees of Freedom, Dimensions of Power*, 145 (1) DAEDELUS 18, 19 (2016) (addressing the ability of online players to predict user behavior and influence it by nudges); RICHARD H. THALER, MISBEHAVING: THE MAKING OF BEHAVIORAL ECONOMICS 341–42 (2015) (addressing how the possibility to shape nudges easily influences their efficiency).

based, massively intermediated environments.[390] By understanding users' cognitive biases, technology can influence their behavior and social dynamics. In practice, intermediaries use nudges to direct user behavior and to influence these users to share more content with a wider audience.[391]

Framing relationships as friendships, selective algorithmic social mirroring of content on user newsfeeds, and explicit and implicit feedbacks constitute nudges. Intermediaries, thus, organize the context in which users make decisions to disseminate information. These influences extend beyond the individual as they generate the social dynamics of dissemination at the macro level of the social network.[392] Yet, every user can choose to disseminate information or avoid dissemination.

Dissemination of content is neither good nor bad in and of itself. While dissemination has many benefits, it can inflict severe harm.[393] The challenge is to diminish dissemination of *offensive* content without chilling information flows in general. Intermediaries should harness the same nudge-based strategies to promote accountability in dissemination and discourage impulsiveness.

Whenever a user of social networks clicks the "publish," "forward," "share," or "re-tweet" buttons, the intermediary should alert him of the risks and consequences of spreading content. The intermediary might turn to users as social actors and raise the following questions: "Are you sure you won't regret sharing this?" or "Could this content cause harm to a third party?" Dissemination could be delayed until the user confirms that he can mindfully share the content. Similarly, intermediaries might inform users of the implications of publishing and sharing content and the information's potential to spread. For example, they might raise questions such as "Do you know that clicking this button exposes the post to your 1,000 friends?" As studies in a related context show, this strategy is likely to cause users to internalize the fact that disseminating content can inflict harm, and thereby promote civility[394]

---

390. *See* COHEN, *supra* note 73, at 180 (explaining that platform-based environments incorporate "choice architecture favoring the decisions that the platform or the application designers want their users to make").

391. For example, framing relationships in social networks as "friendships," feedback mechanisms, and features for sharing content all increase the likelihood an individual will reaching his threshold for disseminating content.

392. *See supra* Section I.D.

393. *See supra* Section I.D (discussing the severe harm of dissemination).

394. *See* THALER & SUNSTEIN, *supra* note 30, at 237 (proposing a similar nudge for civility—a "civility check").

and help them to avoid engaging in regrettable online dissemination.[395]

Some intermediaries already inform users of the potential consequences of sharing content, but this is done only after the fact.[396] A better policy would be to address users before dissemination occurs, in order to prevent automatic intuitive sharing from the start.[397] The idea of using nudges to promote reflective thinking and accountable sharing is starting to gain momentum. For example, Twitter recently began asking people if they are sure that they want to re-tweet a link if they have not accessed the link themselves.[398] Such a nudge can promote reflective thinking before dissemination.

Nudges, therefore, have great potential to mitigate online harm.[399] As technology advances, intermediaries can use these advances to improve nudges. For example, intermediaries can use artificial intelligence and machine learning algorithms to conform the nudge to the content that the user is about to disseminate by identifying specific words in the content.[400] The intermediary can also function as a social actor, instead of simply sending automatic messages, by enhancing user awareness of messages and their possible repercussions.[401]

---

395. *See* Wang et al., supra note 380, at 1318–23, 1331 (arguing that designing mechanisms that nudge users to consider the content and context of their online disclosures are efficient in helping individuals avoid regrettable online disclosures and, thus, enhance their privacy); John et al., *supra* note 380.

396. For example, whenever a user tweets, the intermediary informs him that his tweets will appear in the timelines of his followers. However, this message appears only after dissemination and the "status-quo-bias" reduces the likelihood of *ex post facto* deletion by the user. *See* THALER & SUNSTEIN, *supra* note 30, at 34 (explaining that people have a general tendency to stick with the current situation, this phenomenon is dubbed the "status-quo-bias").

397. *See generally* KAHNEMAN, *supra* note 139, at 19–24 (2011); THALER, s*upra* note 373, at 99 (referring to "the planner" who thinks in the long run versus the "doer" who acts instinctively. Nudging accountability in disseminating content may prevent bypassing deliberation and reduce cognitive biases).

398. See Alex Hern, *Twitter Aims to Limit People Sharing Articles They Have Not Read,* THE GUARDIAN (June 11, 2020), https://www.theguardian.com/technology/2020/jun/11/twitter-aims-to-limit-people-sharing-articles-they-have-not-read.

399. *See* HARTZOG, *supra* note 120.

400. The same strategy intermediaries use to match advertisements to user content also allows intermediaries to use information on user content and interests, in order to tailor nudges to users. On intermediary usage of information to match content and advertisements to users, see Julie E. Cohen, *The Emergent Limbic Media System*, *in* LIFE AND THE LAW IN THE ERA OF DATA-DRIVEN AGENCY 60 (2019) (Mireille Hildebrandt & Kieron O'Hara eds., 2020).

401. WALDMAN, *supra* note 116, at 141 (expounding on social communication of bots that motivate users to release privacy protections, by technological design).

Indeed, social network platforms are starting to use nudges. Facebook cooperates with fact-checking organizations[402] and informs individuals before they share fake news that independent websites and fact checkers have found the information to be false, or at least controversial.[403] Instagram also labels fake news to reduce its dissemination.[404] Twitter followed and added new fact-checking labels to hundreds of tweets. Twitter even used such a label to flag a post by President Donald Trump, rebutting its accuracy, before, during and even after the 2020 election cycle.[405] In another context, Twitter has also started applying warning messages to tweets that contain misleading information about COVID-19.[406] Yet, using this strategy to try to refute false rumors that have already been published and disseminated can backfire. Repeating a falsehood and adding information in order to refute it, or tagging information as false, only exacerbates the information's visibility and increases the likelihood that users will believe it.[407] Tagging falsehoods as such might even lead users to assume that content that was not tagged as false is true, even though it could be completely false.[408]

Moreover, even if exposure to information refuting a falsehood could mitigate its harm, this strategy can be used only regarding content that fact checkers have already deemed false or controversial. This usually happens after content has already spread to a

---

402. *See* Levi, *supra* note 36; BENKLER ET AL., *supra* note 2, at 287 (explaining that the solution of fact checking organizations did not mitigate the problem); *see also* Van Loo, *supra* note 294.

403. Nikhil Sonnad, *This Is Now What Happens When You Try to Post Fake News on Facebook*, QUARTZ (Mar. 19, 2017), http://bit.ly/2LAuOXw.

404. Stephanie Milot, *Instagram Automatically Labels, Hides Fake News*, GEEK.COM (Dec. 17, 2019), https://web.archive.org/web/20191217184642/https://www.geek.com/tech/instagram-automatically-labels-hides-fake-news-1813964/.

405. Kate Conger & Mike Isaac, *Defying Trump, Twitter Doubles Down on Labeling Tweets*, N.Y. TIMES (June 9, 2020), https://www.nytimes.com/2020/05/28/technology/trump-twitter-fact-check.html; *see also* Kim Lyons, *Twitter Flags President Trump's Tweets About Ballot-Counting*, THE VERGE (Nov. 7, 2020), https://www.theverge.com/2020/11/7/21554013/twitter-flags-president-trumps-tweets-votes-counted-election-pennsylvania; *Trump Falsely Claims Victory on Twitter Just Ahead of Biden Win*, THE QUINT (Nov. 7, 2020), www.thequint.com/news/world/won-by-a-lot-president-trump-falsely-declares-victory-on-twitter-again (referring to Trump's misleading tweet: "I won this election by a lot").

406. *See Coronavirus: Twitter Will Label Covid-19 Fake News*, BBC NEWS (May 12, 2020), www.bbc.com/news/technology-52632909.

407. *See generally* DIFONZO & BORDIA, *supra* note 47, at 225 (2007); Pennycook et al., *supra* note 16.

408. *See* David M.J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts & Jonathan L. Zittrain, *The Science of Fake News*, 359 SCI. MAG. 1094–96 (2018); SUNSTEIN, *supra* note 95, at 125 (explaining that trying to refute a falsehood after it has already been published often failed and even exacerbated user commitment to the content of the rumor).

wide audience and caused tremendous harm. Furthermore, initiatives by fact checkers to refute false rumors, defamation, and fake news usually focus on falsehoods about public figures or politicians and neglects to address statements about "ordinary people." In contrast, intermediary nudges for reflective thinking prior to dissemination would apply to all types of false rumors, including rumors about "ordinary people." This type of nudge does not repeat the false rumor or exacerbate its visibility or credibility. Such nudges are neutral to the content that users publish or disseminate and are expected to mitigate the harm of all types of false rumors. Thus, this proposed solution is preferable to current policy.

Companies and policy makers are discovering the potential benefits of nudges.[409] In the related context of privacy and data protection, literature has proposed that choice architects design systems that would generate nudges to enhance informed choices about sharing information and reflective thinking about the privacy implications of sharing.[410] For example, using nudges to encourage users to change the privacy setting of their date of birth on their personal profile and share it with fewer people,[411] or using nudges to reduce the sharing of other private information.[412] Empirical studies have confirmed the efficiency of nudges in privacy protection. For example, researchers found that placing details about the privacy policy of platforms for commerce on search engines pushes users to prefer platforms that maintain higher standards of privacy.[413] Scholars recently discovered the potential of this strategy to combat false rumors, defamation, and fake news.[414] Yet, literature has not held a comprehensive discussion on this related context.

Due to the potential of nudge-based strategies to slow the dissemination of such content, intermediaries should apply this solution to mitigate irresponsible dissemination of falsehoods. At this junction, the question is how to incentivize intermediaries to incorporate nudges for accountability in dissemination. More broad-

---

409. SUNSTEIN & HASTIE, *supra* note 97, at 107, 114 (referring to the strategy of asking simple questions to promote critical thinking, change perspectives and avoid information and reputation cascades).

410. HARTZOG, *supra* note 120, at 215–26; Alessandro Acquisti, Laura Brandimarte & George Loewenstein, *Privacy and Human Behavior in the Age of Information*, 347 SCI. MAG. 509, 511 (2015).

411. *See* Egelman et al., *supra* note 355.

412. Wang et al., *supra* note 380, at 1334 (found that "privacy nudges can potentially be a powerful mechanism to help some people avoid unintended disclosures").

413. *See, e.g.*, Acquisiti et al., *supra* note 410.

414. *See* Nancy S. Kim, *Web Site Proprietorship and Online Harassment*, UTAH L. REV. 993, 1014 (2009); Woodrow Hartzog & Evan Selinger, *Increasing the Transaction Costs of Harassment*, B.U. L. REV. ANNEX 47 (2015).

ly, should intermediaries adopt this solution voluntarily, or should the law obligate them to nudge?

At first glance, it would appear that intermediaries are not likely to adopt solutions that aspire to reduce dissemination because they currently profit from it and even use nudges to promote dissemination.[415] Arguably, legal regulation is preferable to incentives for voluntary adoption of nudges, meaning the government should force intermediaries to nudge. A closer look, however, reveals that this is not always the case. Here, it is important to differentiate between the dissemination of content in general and the spreading of falsehoods. Indeed, in most cases, intermediaries have no incentive to self-regulate dissemination because they profit from it. With respect to falsehoods, though, the situation is different. Many intermediaries are concerned by "the potential business, moral, and instrumental costs" of falsehoods and see these falsehoods "as a potential threat to [their] profit[s]."[416] Other intermediaries self-regulate falsehoods due to a sense of corporate responsibility,[417] to enhance their social standing,[418] or as a preventive measure to shy away from murky legal areas and diminish the likelihood of claims against them. Due to the potential benefits, many intermediaries already self-regulate dissemination of falsehoods.[419] Nudges for accountability focus on offensive content and do not purport to reduce dissemination generally. Therefore, intermediaries might have an intrinsic motivation to adopt this solution.[420]

In addition, other incentives can be created. For instance, fiduciary intermediaries could grant a stamp of approval (such as a trust mark) to intermediaries that promote user accountability.[421]

---

415. *See supra* Section I.C and accompanying notes (describing strategies that intermediaries utilize to enhance social sharing).

416. Lavi, *supra* note 54, at 937; *see also* Thomas E. Kadri, *Networks of Empathy*, 4 UTAH L. REV. 1088 (2020) ("[P]erhaps because the design will create goodwill and help to retain a company's customers or because it will entice more people (including victims) to start using their products.").

417. *See* Klonick, *supra* note 294, at 1627; Citron & Norton, *supra* note 108, at 1455–56 (noting that voluntary regulation can be justified by doctrines of corporate law, which allow managers to consider public interests).

418. *See* BAMBERGER & MULLIGAN; *supra* note 354, at 35 (referring to a "social license to operate").

419. *See* Citron & Norton, *supra* note 108, at 1454 (discussing MySpace as an example of voluntary regulation by a mediator due to economic considerations and voluntarily removal of offensive content in order to attract children). Yet, in most cases, self-regulation focuses on content moderation and *ex ante* measurement and does not fulfil the promise of *ex ante* regulation by design.

420. *See* Klonick, *supra* note 294, at 1606–07, 1626–30 (explaining that despite immunity of intermediaries from liability in the United States, intermediaries voluntarily self-regulate due to corporate responsibility and to create a more pleasant environment for users that would make their platform more attractive and enhance profits).

421. In our context, ISOC (Internet Society) can function as a fiduciary intermediary. *See Frequently Asked Questions,* INTERNET SOCIETY (May 2019), http://www.isoc.org/isoc/general.

This might improve their image, attract more users, and therefore enhance advertising profits and contribute to the intrinsic motivation to prevent dissemination of offensive content. Of course, illicit intermediaries exist. Such intermediaries encourage distribution of offensive content and profit particularly from its publication and dissemination.[422] These intermediaries lack any economic or moral incentive to adopt the proposed solution. In such cases, legal regulation is also not an efficient solution. As studies in related contexts prove, when a party opposes the policy behind the nudge, obligating it to adopt the policy is inefficient and the party is likely to nudge unconvincingly, by steering users away from it, using dark patterns, and manipulating nudge-forcing rules.[423] Thus, mandates impose expensive costs without reaping any benefits.

In summary, nudging to promote reflective thinking is not a perfect solution. It can, however, reduce impulsive dissemination by neutrals and skeptics, discourage impulsive postings, and slow down the dissemination of falsehoods. It improves upon current policy, which does not disincentivize these types of propagators. Incentivizing intermediaries to adopt this solution voluntarily is superior to legal regulation, since legal regulation is less flexible and might infringe on the intermediary's freedom to conduct their business. The lack of flexibility would also likely have negative consequences for efficiency and innovation. When the intermediary

---

Trust marks are defined as "[e]lectronic labels or visual representations indicating that an e-merchant has demonstrated its conformity to standards regarding security, privacy, and business practice." *See* EUROPEAN CONSUMER CTRS.'S NETWORK, TRUST MARKS REPORT 2013: "CAN I TRUST THE TRUST MARK?" (2013), https://ec.europa.eu/info/sites/info/files/trust_mark_report_2013_en.pdf (discussing how trust marks can be used to protect consumers).

422. The business models of these intermediaries are based on dissemination of sensational rumors, which may in turn draw more users. *See* Jones v. Dirty World Entertainment Recordings LLC, 755 F.3d 398 (6th Cir. 2014). In addition, some intermediaries encourage users to share negative reviews and directly profit from defamation through Corporate Advocacy Programs, which purport to assist in resolving the posted complaints. They charge victims money for removing the offensive content. *See, e.g.*, Hy Cite Corp. v. Badbusinessbureau.com, L.L.C., 418 F. Supp. 2d 1142, 1149 (D. Ariz. 2005); Vo Group v. Opinion Corp., No. 8758/11 (N.Y. Sup. Ct. May 22, 2012); *see also* Lavi, *supra* note 53; Kim, *supra* note 414, at 1045; Ann Bartow, *Internet Defamation as Profit Center: The Monetization of Online Harassmen*t, 32 HARV. J.L. & GENDER 384 (2009).

423. For example, the intermediary can use small letters or cumbersome language. *See* Lauren E. Willis, *When Nudges Fail: Slippery Defaults*, 80 U. CHI. L. REV. 1155, 1200–01 (2013) (discussing instances in which a party stands to lose revenue due to a nudge—such as a bank and automatic enrollment in anti-overdraft programs—and thus makes an effort to steer users away from its influence). On dark patterns in a related context, see Ari Ezra Waldman, *Power, Process, and Automated Decision-Making*, 88 FORDHAM L. REV. 613, 619–20 (2019); Ari Ezra Waldman, *Cognitive Biases, Dark Patterns, and the "Privacy Paradox,"* CURRENT ISSUES PSYCH. (forthcoming 2020); Jamie Luguri & Lior Jacob Strahilevitz, *Shining a Light on Dark Patterns*, 12 J. LEGAL ANALYSIS (forthcoming 2020) (expanding on dark patterns in the context of privacy and data protection).

lacks intrinsic motivation to adopt this solution, it will likely bypass regulation. Therefore, a voluntary solution is preferred.

C. *Nudges for Accountable Dissemination of Information: Addressing Limitations and Objections*

The solution of nudges for accountable dissemination is promising; however, there are certainly several potential objections to this framework that must be addressed. First, general scholarly criticism of nudges should be addressed. Nudges have raised many controversies, objections, and ethical concerns in scholarly work.[424] It has been argued that nudges are not libertarian paternalism but actual paternalism in disguise;[425] they manipulate choices and should be constrained.[426] In light of this criticism, one could argue that it is inappropriate to adopt such a controversial nudge-based solution.

The general controversy regarding nudges is beyond the scope of this Article. Indeed, some nudges can be manipulative and unethical.[427] However, as far as the proposed nudges are concerned, most objections are marginal because, in this context, nudges are not manipulative and can even promote individual autonomy. Thus, those who object to nudges in general may agree that the nudges proposed here should not be constrained. Intermediaries will influence decisions to share content by raising questions that provoke users' reflective thinking and, in turn, prevent impulsive dissemination. As empirical studies have shown, individuals sup-

---

424. *See* Cass R. Sunstein, *The Ethics of Nudging* (Harv. L. Sch. Discussion Paper, No. 806, 2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2526341 (weighing arguments for and against nudge-based strategies).

425. *See* Thaler et al., *supra* note 374 (arguing that the idea of libertarian paternalism is both possible and legitimate for private and public institutions); *see also* Henry Farrell & Cosma Shalizi, *'Nudge' Policies Are Another Name for Coercion*, NEW SCIENTIST (Nov. 2, 2011), https://www.newscientist.com/article/mg21228376-500-nudge-policies-are-another-name-for-coercion (arguing that nudges are paternalistic coercion).

426. *See* THALER & SUNSTEIN, *supra* note 30, at 237 (referring to objections against nudge-based strategies and emphasizing that there is no completely "neutral" design). However, other scholars differentiate between "a given context that *accidentally* influences behavior and a choice architect *who intentionally* tries to alter behavior by fiddling with contexts." *See, e.g.*, Guldborg Hansen & Andreas Jespersen, *Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behavior Change in Public Policy*, EUR. J. RISK REG. 3, 10(2013); T. M. Wilkinson, *Nudging and Manipulation*, 61 POL. STUD. 341 (2013).

427. There are nudges that apply to intuitive thinking and might be manipulative and objectionable. *See* SUNSTEIN, *supra* note 371, at 17, 82; Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, 4 GEO. L. TECH. REV. 1, 19 (2019) (discussing the relationship between nudges and manipulation in an online context). Yet, nudges can be transparent and promote reflective thinking. *See* SUNSTEIN, *supra* note 375, at 17.

port this type of nudge, which enhances their self-control[428] and seeks to prevent impulsive dissemination.[429] These nudges do not aim to affect subconscious or unconscious processing of information. Therefore, they should not be considered manipulation or an objectionable interference with autonomy.[430]

The proposed nudges may also counter the influence of choice architecture, which promotes dissemination in general, appeals to intuitive thinking, and leads individuals to regret sharing content after the fact.[431] These nudges are transparent and help individuals correct mental shortcuts to achieve legitimate objectives. In fact, they promote autonomy and self-governance,[432] and do not insult individual dignity.[433] This type of nudge respects individual goals and promotes public welfare.[434] Hence, the general criticism against nudges does not undermine the proposed solution.[435]

Second, it can be argued that the nudge solution is inefficient. Even if most intermediaries adopt the nudges for accountable dissemination, nudges will influence only some propagators and disseminators. Indeed, nudges are expected to be useful in dissuading altruistic propagators, who believe the information they intend to publish is true, from publishing the information. Nudges are also likely to be useful in dissuading neutrals, who have no inclination in favor of or against the information and skeptics, who have higher thresholds for the dissemination of content, and have the potential to break the "follow the herd" mentality. However, nudges will not stop malicious propagators who publish offensive content

---

428. *See generally* Sunstein, *supra* note 378 at 184 (discussing how there is greater support for nudges that appeal to a person's capacity for reflective and deliberate choice than for nudges that seem to affect the subconscious); Cass R. Sunstein, Lucia A. Reisch & Micha Kaiser, *Trusting Nudges? Lessons from an International Survey*, 26 J. EUR. PUB. POL'Y 1417, 1421 (2018).

429. *See generally* KAHNEMAN, *supra* note 139.

430. *See* Cass R. Sunstein, *Fifty Shades of Manipulation*, 1 J. MKTG. BEHAV. 213, 217 (2016); CASS SUNSTEIN, THE ETHICS OF INFLUENCE: GOVERNMENT IN THE AGE OF BEHAVIORAL SCIENCE 53–54 (explaining that nudges that promote welfare, autonomy, dignity, and self-governance are ethical).

431. For instance, identifying a relationship on Facebook as a "friendship" appeals to intuitive thinking and enhanced sharing. *See supra* Section I.C and accompanying notes; Grimmelmann, *supra* note 87, at 1179–81. On the potential of defaults that are considered nudges to promote autonomy, see generally Cass Sunstein, *Autonomy by Default*, 11 AM. J. BIOETHICS 1 (2016); and SUNSTEIN, *supra* note 430, which refers to educative nudges that lead individuals to make better choices for themselves.

432. *Cf.* SUNSTEIN, *supra* note 430, at 72, 74–77; Susser et al., *supra* note 427, at 21 (explaining that lies are manipulative but that the proposed nudges aim to correct cognitive biases and mitigate the harm of lies and falsehoods, and thus they are not manipulative).

433. *E.g.*, SUNSTEIN, *supra* note 430, at 60 (giving the example of the GPS that helps individuals navigate without insulting anyone's dignity).

434. *See id.* at 53–54 (discussing how governments use nudges to increase welfare).

435. *See id.* at 53–77 (explaining that nudges that promote welfare, autonomy, and self-governance are not manipulative).

simply to inflict injury.[436] They are not likely to dissuade narrowly self-interested propagators and have very little influence on generally self-interested propagators from publishing falsehoods. In addition, receptives, who have a low threshold for accepting and adopting information, may continue to spread false rumors, defamation, and fake news despite nudges.[437]

Moreover, nudges have no influence on bots that disseminate rumors and operate by technological code. These Artificial Intelligence entities are usually operated by narrowly self-interested individuals[438] who program the code to publish and echo specific falsehoods.[439] In order to allow falsehoods to spread, the operators of bots are likely to program the algorithm to automatically signal that they "agree" to publish or share the content despite the intermediary's alerts.[440]

Indeed, nudges are likely to influence only part of the network, and yet, this solution can still reduce and slow down dissemination of false rumors, defamation, and fake news within large parts of a network. Thus, fewer informational cascades would be expected. As explained in Part I above, the more times a post is shared, the more individuals reach their threshold of belief and disseminate it further. Although nudges influence only part of the network, they can reduce the publishing and sharing of falsehoods and can slow down their dissemination, therefore, minimizing the gravity of the harm falsehoods inflict.

Truly, nudges cannot influence automatic algorithms and bots that disseminate and share falsehoods. Such algorithms pose challenges to the protection of reputations and the public interest,[441] and it might be advisable to adopt additional mechanisms or spe-

---

436. *See generally* SUNSTEIN, *supra* note 41, at 13.

437. Due to the receptive's prior disposition in favor of the rumor, they have a low-level threshold. Therefore, a nudge may not suffice to dissuade them from spreading it.

438. On the motivations to publish and spread rumors, see *supra* Section I.A.1.

439. Engineers who serve companies and stakeholders control the parameters at the base of the algorithms *ex ante*. *See* Jack M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L.J. 1217, 1224 (2017) ("When we criticize algorithms, we are really criticizing the programming, or the data, or their interaction. But equally important, we are also criticizing the use to which they are being put by the humans who programmed the algorithms, collected the data, or employed the algorithms and the data to perform particular tasks.").

440. *Massive Networks of Fake Accounts Found on Twitter*, BBC NEWS (Jan. 24, 2017), https://bbc.com/news/technology-38724082. For more information on Twitter bots during the 2016 U.S. elections, see Philip N. Howard, Samuel Woolley & Ryan Calo, *Algorithms, Bots, and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration*, 15 J. INFO. TECH. & POL. 81 (2018).

441. *See* Lazer et al., *supra* note 408, at 1095 (discussing how dissemination of fake news can erode individuals' trust in reputable news outlets and make it harder for people to obtain truthful information).

cific regulations pertaining to bots.[442] Many intermediaries have intrinsic incentives to narrow down the activity of bots on their platforms because in many cases bots fail to comply with the platform's terms of service and distort the public discourse. For example, many intermediaries already act to reduce activities of fake bots by using algorithms.[443] Moreover, bots are less likely to be involved in all types of dissemination of falsehoods. Rather, they focus on commercial entities, public figures, and public representatives.[444] Dissemination of falsehoods about such entities has implications beyond the direct parties. In such cases, it is likely that local and state law enforcement will get involved and invest resources in identifying the parties that are directly responsible for the dissemination.[445] Civil society organizations are also likely to be involved in detecting bots[446] and to function as watchdogs who report bots to

442. For example, one proposal is to impose a duty on paid influencers to carry a disclaimer informing users that they are paid, and the source of payment. *See* BENKLER ET AL., *supra* note 2, at 371–75; Honest Ads Act, H.R. 4077, 115th Cong. § 1989 (2017) (advocating for disclosure obligations by those who are paid to advertise things to the public). The trigger for the Honest Ads Act was Russian intervention in the U.S. 2016 elections and the need to ensure that electioneering communities are not funded by foreign nationals. Ellen P. Goodman & Lyndsey Wajert, *The Honest Ads Act Won't End Social Media Disinformation, but It's a Start* (Nov. 2, 2017), (unpublished manuscript), papers.ssrn.com/sol3/papers.cfm?abstract_id=3064451.

443. In the context of extremist content, Facebook is using artificial intelligence (AI) and machine learning to combat harmful content more efficiently. Platforms are expected to find better ways to combat fake news with the development of this technology. *See, e.g.*, Julia Fioretti, *Pressured in Europe, Facebook Details Removal of Terrorism Content*, REUTERS (June 15, 2017), https://www.euractiv.com/section/politics/news/pressured-in-europe-facebook-details-removal-of-terrorism-content/. It should be noted that bots are becoming more sophisticated and are learning to obscure indication as automated entities, in an arms race with intermediaries, which are improving their strategies to detect them. For example, the intermediary can reduce the participation of nonhuman software by using CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) and some technologies allow intermediaries to detect and remove "likes" that are generated by "automated software programs, malware, and hacked accounts." *See* Lucille M. Ponte, *Mad Man Posting as Ordinary Consumer: The Essential Role of Self-Regulation and Industry Ethics on Decreasing Deceptive Online Consumer Rating and Reviews*, 12 J. MARSHALL REV. INTELL. PROP. L. REV. 462, 504 (2013); FINN BRUNTON, HELEN NISSENBAUM: OBFUSCATION – A USER'S GUIDE FOR PRIVACY AND PROTEST 40 (2015).

444. *See* Madeline Lamo & Ryan Calo, *Regulating Bot Speech,* 66 UCLA L. Rev. 988, 995–1002 (2019) (outlining a taxonomy of three main types of bots: commercial , political and creative). Arguably, a person is less likely to invest efforts and code algorithms to echo a negative falsehood on private 'ordinary' people. Rather, it seems more reasonable that bot operators focus on spreading information on commercial entities and public figures. It can be assumed that creative bots are less likely to spread defamation on private people, because they operate for creative purposes.

445. Law enforcers are already investigating the activity of bots that influence elections. *See, e.g., Report: FBI Investigating Russian Operatives Using Bots to Spread Stories from Breitbart, RT, Info Wars,* MEDIA MATTERS (Mar. 20, 2017), https://www.mediamatters.org/breitbart-news/report-fbi-investigating-russian-operatives-using-bots-spread-stories-breitbart-rt.

446. *See, e.g.*, Press Release, Oxford Internet Institute, Oxford Experts Launch New Online Tool to Help Fight Disinformation (Dec. 10, 2019), https://www.oii.ox.ac.uk/news/releases/oxford-experts-launch-new-online-tool-to-help-fight-disinformation/ ("Researchers at the Oxford Internet Institute, University of Oxford, have launched 'The ComProp

social network operators and law enforcers.[447] Thus, market forces and the law are likely to narrow the gap in self-regulation by nudges and reduce dissemination of falsehoods by bots.

Third, the efficiency of nudges can decrease over time. Arguably, users will get used to nudges, learn to ignore them, and click automatically on the button that allows them to publish and disseminate information without paying any attention to the intermediary's alerts.

This does not undermine the proposed strategy. The use of nudges can still slow down dissemination of falsehoods relative to dissemination on platforms without nudges.[448] In addition, the costs of nudges in the digital ecosystem are lower than in a brick-and-mortar infrastructure, allowing their benefits to outweigh their costs. Furthermore, technology advancements could preserve the efficiency of nudges.[449] The intermediary could learn to identify users that consistently ignore nudges and immediately click and share and develop stronger nudges that are more likely to influence them.[450] For example, the intermediary might increase the diversity of nudges, personalize them, and make them more relevant to the specific content that the user intends to publish or share. Thus, the intermediary might be able to preserve the efficiency of the nudge over time.

Fourth, nudges can result in a chilling effect and suppress the motivation to publish and disseminate legitimate expression, in particular among minorities that might avoid expressing ideas that are outside of the consensus. This argument is valid, yet the nudge strategy brings about a better balance between freedom of speech, dignity, and public interest in an era of vast digital dissemination. Although nudges have the potential to chill speech, they do not prohibit anyone from speaking and the choice to speak remains in the hands of every individual. Nudges might cause minorities and other disadvantaged groups in society to self-censor legitimate expressions that are outside of the consensus, but nudges can also

---

Navigator,' a new online resource guide which aims to help civil society groups better understand and respond to the problem of disinformation.").

447. *See, e.g.*, Bergman, *supra* note 42.

448. *See, e.g.*, THALER & SUNSTEIN, *supra* note 30, at 237–38 (proposing the use of nudges to enhance civility among individuals online).

449. *Cf.* Tal Z. Zarsky, *Privacy and Manipulation in the Digital Age*, 20 THEORETICAL INQUIRIES L. 157, 184 (2019) (suggesting that, in the context of intermediaries and advertiser influences, as consumers get used to strategies of influence and learn to ignore them, companies may develop new ways of influencing consumers).

450. *See, e.g.*, Adrian L. Jessup Scheneider & T.C. Nicholas Graham, *Pushing Without Breaking: Nudging Exergame Players While Maintaining Immersion*, 2015 I.E.E.E. GAMES ENT. MEDIA CONF., Oct. 14–16, 2015, at 1, 1–8 (2015) (discussing the use of progressively severe nudges in response to players who ignore nudges in the related context of exergames).

mitigate potential harm to minorities and disadvantaged groups when falsehoods and defamation are disseminated against them. Minorities and disadvantaged groups suffer from harmful expression more than other sectors of the population.[451] Thus, nudges can narrow social gaps and improve protection of the human rights of the disadvantaged as potential victims.

D. *Efficient Removal* Ex Post Facto

A common remedy for dissemination of falsehoods is *ex post facto* removal. This remedy, however, does not solve the problem of disseminating falsehoods. A victim whose name has been tainted would have to indicate each and every virtual setting of a falsehood or defamation in order to completely remove the content and ensure it would not be found and widely disseminated again. Civil society organizations might also find it difficult to report on every location of fake news and protect the public interest. Due to the vast dissemination of content within social networks, complete removal may be impossible.

This problem calls for a solution that enables the removal of falsehoods and defamatory remarks from all settings in which they reside. Unlike the previous solution, which focused on behavioral influences of architecture, this solution focuses on the code itself, the constraints it constitutes, and the possibilities it affords.[452] One possibility is to design the code to allow the dissemination of a message to only a limited number of recipients, thereby slowing down the dissemination of falsehoods, and even preventing them from reaching the tipping point and going viral.[453] Yet this solution could hinder the dissemination of useful, newsworthy information. A preferable solution would be efficient removal *ex post facto*.[454] Intermediaries that integrate features that allow for the sharing of content within their platforms should also allow the efficient removal of harmful content from any profiles and locations to which the content was disseminated by the removal of the original

---

451. *See, e.g.*, Thomas H. Koenig & Michael L. Rustad, *Digital Scarlet Letters: Social Media Stigmatization of the Poor and What Can Be Done*, 93 NEB. L REV. 592, 596–601 (2015).

452. *See* LESSIG, *supra* note 342, at 123–25.

453. *See, e.g.*, Jacob Kastrenakes, *WhatsApp Limits Message Forwarding in Fight Against Misinformation*, THE VERGE (Jan. 21, 2019), https://www.theverge.com/2019/1/21/18191455/whatsapp-forwarding-limit-five-messages-misinformation-battle (discussing WhatsApp's limit on the number of times a message can be forwarded).

454. For a previous article in which I discuss this solution, see Lavi, *supra* note 32.

post.[455] Intermediaries can do so, by integrating in the code of their platforms features such as an "embedded link."[456]

"Embedded links allow users to import external web content and present it in their profiles."[457] Removal of an original post with an embedded link would result in deletion of all replications disseminated by the "share" button upon complaint from the victim, a judicial decision, or flagging practices. As every replication of the post that was created by the share button is connected to the original publication, even reports on a replication are likely to lead to flagging the original post as inappropriate. Alternatively, intermediaries could use "technology that allows data tethering, which changes the shared content according to the source."[458] The intermediary can integrate these features in the code, architecture, or protocol of their platforms at the stage of design. This technology is in fact used today;[459] "however, choice architecture is value-laden and reflects a particular set of preferences that should not be taken for granted."[460] The values behind technology can influence the way it is used.[461] The design of the platform and code can make information more visible or more obscure.[462] It can create an incentive to upload and share more content almost automatically[463] or, by contrast, encourage reflective thinking before sharing posts. Similarly, design can either make it easy to share content at the click of a button or do exactly the opposite by increasing the costs of dissemination. For example, intermediaries can make it difficult to share content by designing architecture that allows dissemination only by copying-and-pasting of content to every recipient. Alternatively, intermediaries can limit the number of recipients with

---

455. *Id.* at 2670.

456. Lavi, *supra* note 32, at 2670–71. For further discussion of the context of intellectual property, see Toby Headdon, *An Epilogue to Svensson: The Same Old New Public and the Worms that Didn't Turn*, 9 J. INTELL. PROP. L. & PRAC. 662 (2014).

457. Lavi, *supra* note 32, at 2671.

458. Lavi, *supra* note 32, at 2671 (citing JONES, *supra* note 243 at 187 (explaining that technology can allow every copied piece of data to be tethered to its master copy)).

459. *See supra* note 31 and accompanying text.

460. Lavi, *supra* note 32, at 2671.

461. *See generally* Mulligan & Bamberger, *supra* note 110, at 701 (explaining that the design of technology is an effective mean of control); HARTZOG, *supra* note 120, at 44–45 (explaining that design is political; it can suppress the dissemination of personal information and promote privacy, or promote values of freedom of information).

462. For example, the design and architecture of the platform can make it difficult to find information on users as it can make information more obscure, to increase the costs of finding information on users and enhance privacy. In contrast, design can enhance access to information, reduce the costs of finding it, and infringe on user privacy. Design can also include dark patterns and obscure objectionable terms of service. *See id.* at 272.

463. *See* FRISCHMANN & SELINGER, *supra* note 130, at 11 (explaining that "[I]t's rapidly becoming easier to design technologies that nudge us to go on auto-pilot and accept the cheap pleasure that comes from minimal thinking . . . .").

whom the user can share information with the click of a button or allow efficient removal of the content shared *ex post*. By designing efficient removal mechanisms, removal of the original content can lead to widespread removal from all profiles. Every intermediary designs its platform in a way that promotes its own objectives.[464] Some intermediaries tune their sharing mechanisms so that users share a link to the original post. For instance, "a click on the 'share' button on Facebook or the 're-tweet' button on Twitter links to the original content and embeds the shared content into the profiles of the disseminator and his friends."[465] Yet, there are different business models and attitudes regarding content moderation and removal of offensive content. Alongside intermediaries with an intrinsic motivation to moderate content out of a sense of social responsibility,[466] or in order to enhance their social standing and attract a greater audience,[467] there are platforms that profit from offensive sensational content, such as gossip websites or extreme racist alt-right websites. Such websites have no intrinsic incentive to design technology for efficient removal of content.[468]

Moreover, even mainstream media giants do not share a uniform policy regarding removal of offensive content. For example, in the related context of incitement to terrorism, "Twitter used to take a laissez-faire approach to terrorist content and avoided removing it even [when] it was made aware of the content,"[469] while Facebook made efforts to remove the content upon knowledge.[470] Twitter changed its policy only when regulation of extremist speech became a real possibility.[471] Similarly, different intermediar-

---

464. *See* Lavi, *supra* note 32, at 2671. Designs, for example, can promote privacy protection or infringe on it. Intermediaries use the design by including dark patterns and obscuring objectionable terms of service. *See* HARTZOG, *supra* note 120, at 272; Nancy Kim, *Website Design and Liability*, 52 JURIMETRICS 383, 402–03 (2012).

465. Lavi, *supra* note 31, at 2671.

466. *See* Klonick, *supra* note 294.

467. Citron & Norton, *supra* note 107, at 1453–57 (discussing MySpace as an example of an intermediary's voluntary regulation due to economic considerations and voluntarily removal of offensive content in order to attract children).

468. *See, e.g.*, Emma Grey Ellis, *Gab, the Alt-Right's Very Own Twitter, Is the Ultimate Filter Bubble*, WIRED (Sept. 14, 2016), https://www.wired.com/2016/09/gab-alt-rights-twitter-ultimate-filter-bubble/; MARANTZ, *supra* note 39, at 5–6; Lavi, *supra* note 78, at 15.

469. Lavi, *supra* note 138, at 498.

470. Nina I. Brown, *Fight Terror, Not Twitter: Insulating Social Media from Material Support Claims*, 37 LOY. LA. ENT. L. REV. 1, 10 (2016).

471. *See* Michelle Roter, *With Great Power Comes Great Responsibility: Imposing a Duty to Take down Terrorist Incitement on Social Media*, 45 HOFSTRA L. REV. 1379, 1391–1392, 1399 (2017) (referring to an official statement from the White House encouraging social media platforms to block more terrorists from using their services and discussing Twitter's policy). For information on Twitter's policy, see *Combating Violent Extremism*, TWITTERBLOG (Feb. 5, 2015), blog.twitter.com/en_us/a/2016/combating-violent-extremism.html. *See generally* Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE

ies have different attitudes towards moderation of posts that incite violence. For example, while Twitter hid President Trump's May 2020 tweet "glorifying violence," Facebook avoided hiding or removing it.[472] Thus, legislative changes regarding efficient removal mechanisms can provide an incentive for removal even when the intermediary lacks intrinsic motivation to mitigate the harm of offensive content.

Unlike the previous solution of nudges, which focused on arranging the decision-making context, this solution applies directly to the code itself and the possibilities it affords and focuses on the result (efficient removal). Therefore, there are fewer opportunities for intermediaries that lack intrinsic motivation to evade implementation. In this case, a combination of legal and technological measures is preferable to voluntary self-regulation.

From a normative perspective, as technologies advance and the moderation and influence of intermediaries on the flow of information become a fundamental aspect of any platform, it is high time to rethink their legal obligations. Scholars have proposed that the law should refine the immunity regime granted to intermediaries in § 230.[473] Some have asserted that the law should adopt a version of a "notice-and-takedown" regime for social network platforms.[474] Yet, a condition should be added to this safe haven. Accordingly, the safe haven should apply only upon efficient removal of offensive content shared via publish, share, or re-tweet buttons. Conditioning the safe haven on efficient removal measures can mitigate injury to victims since they would not need to point out every location where the falsehood was shared in order to remove it. Instead, the victim could send a "notice-and-takedown" notification to the intermediary regarding the original post only. Taking down the original post would lead to removal of all replications that were created using the share button.

Efficient "removal of offensive content might [also] be achieved even without amending [§] 230,"[475] by adopting other formal regu-

---

DAME L. REV. 1035, 1037 (2018) (referring to changes in policies of social media platforms adopted to stave off the threat of European regulation).

472. Bond, *supra* note 307.

473. 47 U.S.C. § 230. For scholarly proposals to narrow down immunity, see Citron & Wittes, *supra* note 174, at 420; Danielle Keats Citron & Benjamin Wittes, *The Problem Isn't Just Backpage, Revisiting Section 230 Immunity*, 2 GEO. L. TECH. REV. 453, 470–71 (2018); Danielle Keats Citron, *Section 230's Challenge to Civil Rights and Civil Liberties*, KNIGHT FIRST AMENDMENT INST. COLUM. U., 4–5 (Apr. 6, 2018); and Citron, *supra* note 114, at 1074.

474. See *supra* Part II for a discussion on this regime. *See also* Citron & Wittes, *supra* note 473, at 470 ("Sustained failure to remove an account despite repeated notifications . . . might well strip the company of immunity in a specific case."); Lavi, *supra* note 52, at 930–31 (arguing that a version of notice-and-takedown regime is optimal for regulating secondary liability of intermediaries on social network platforms).

475. Lavi, *supra* note 32, at 2669.

lations. Such regulations might impose an obligation to design efficient mechanisms to remove replications of unwanted posts along with the original post. In many cases, the law in the United States allows victims of alleged defamation to file legal suits against the person that published the original post.[476] A defamation suit is possible if the plaintiff uses his real name and can be identified.[477] Many website platforms allow the original publisher of posts or comments to remove them. The publisher may then remove his posts or comments at the request of the victim, following a direct lawsuit filed by the victim, or if a court orders the defendant to do so.[478] Such regulation would lead to removal of all replications along with the publisher's removal of the original post.

Moreover, in spite of the immunity outlined by § 230, "some courts may order the intermediary to remove the offensive statements" following a ruling that the statements were defamatory.[479] A lower court in California already extended an order to remove defamatory online reviews to the intermediary. Even though the Supreme Court of California reversed this ruling, extending an order to remove harmful content to third-party platforms can be adopted in other states.[480] Moreover, the practice of enjoining non-liable platforms to defamation lawsuits can develop.[481] This may allow efficient removal of falsehoods when the publisher does not appear at a court hearing or refuses to take statements down.

"This solution mitigates [the] harm of the victim, enhances the publisher's control over content, and allows him to remove the content *ex post facto*."[482] It reduces administrative costs and promotes efficiency while preserving freedom of expression and avoid-

---

476. *See, e.g.*, Boulger v. Woods, 917 F.3d 471, 475 (6th Cir. 2019) (dismissing claim because the defendant added question marks to his defamatory tweets, making it not actionable); *see also* Lavi, *supra* note 47, at 2669; Hunt, *supra* note 19, at 560–62 (reviewing defamation lawsuits regarding libelous statements posted on Twitter).

477. Many social network platforms, including Facebook, require users to construct a profile that reflects their real identity ("real-name policy") and use their offline names when interacting within the platform. *See Statement of Rights and Responsibilities*, FACEBOOK, https://www.facebook.com/terms (last modified Jan. 30, 2015); Lavi, *supra* note 32, at 2669.

478. Lavi, *supra* note 32, at 2669.

479. *Id.* at 2670.

480. *Id.; see, e.g.*, Hassel v. Bird, 203 Cal. Rptr. 3d 203 (Ct. App. 2016) (granting default judgment and injunctive relief to a lawyer who sued Yelp for a defamatory review), *rev'd*, 420 P.3d 776 (Cal. 2018); *see also* Eric Goldman, *The California Supreme Court Didn't Ruin Section 230 (Today)*—Hassell v. Bird, TECH. & MKTG. L. BLOG (July 2, 2018), https://blog.ericgoldman.org/archives/2018/07/the-california-supreme-court-didnt-ruin-section-230-today-hassell-v-bird.htm.

481. For a proposal of enjoining non liable platforms and allowing courts to grant injunctions of removal directed at the platforms, see Maayan Perel, *Enjoining Non-Liable Platforms*, 34 HARV. J.L. & TECH. 1, 30–41 (2020).

482. Lavi, *supra* note 32, at 2672.

ing disproportionate infringement on the intermediary's right to conduct a business.

The concept of a legal obligation for intermediaries to implement efficient removal technology at the design stage of the platform, together with defamation suits to remove the content, might mitigate harm mainly in cases of negative falsehoods against private individuals because the standard of liability is lower relative to the standard in defamation suits of public-figures.[483] Yet, in cases of severe fake news against public figures, state and local law enforcement authorities could be involved and might order the removal of fake news.[484] In such cases, an efficient removal mechanism can increase the likelihood of achieving better results in the removal of fake news.

E. *Efficient Removal* Ex Post Facto*: Addressing Limitations and Objections*

First, it can be argued that the solution of efficient removal of shared content is insufficient because only content that was disseminated using the platform's sharing feature buttons would be removed.[485] Users would still be able to copy false or defamatory posts and paste them elsewhere on the platform without using any of the platform's built-in sharing features.[486] These replications would remain on the platform even if the original post is removed, thus undermining the efficiency of the proposed solution.

This argument underlines a genuine weakness of the solution. In contrast to clicking a button, however, copying and pasting content elsewhere on the platform is cumbersome, and such a manual sharing is not automatic. Thus, while the efficient removal of content *ex post facto* is not a perfect solution, it will likely reduce the reputational harm of victims of falsehoods.[487]

---

483. In *New York Times Co. v. Sullivan*, 376 U.S. 254, 279–80 (1964), the U.S. Supreme court held that malice must be proven rather than presumed in cases involving the alleged defamation of public officials. In contrast, the Court has held that a private-figure plaintiff must show, at a minimum, that the defendant was negligent in verifying the allegedly defamatory false claim; thereby it is much more difficult for a public figure to win a defamation lawsuit. *See* Gertz v. Robert Welch, 418 U.S. 323, 347 (1974).

484. States (or supranational entities like the E.U.) are trying to regulate key players that shape the internet in order to influence them to design infrastructure to control the content that users post and disseminate. This is already happening in the context of terrorist content and fake news and incitement to terror. *See* Michal Lavi, *supra* note 141, at 506; Balkin, *supra* note 332.

485. *See* Lavi, *supra* note 32, at 2673.

486. *See id.*

487. *See id.* at 2670–71.

Moreover, with the advance of technology, market forces are likely to bridge the gap in regulation.[488] Social media giants like Facebook and Twitter can cooperate and share unique digital fingerprints that are assigned to videos or photos containing extreme defamation, false rumors, or unambiguously fake news and that have been removed from one of their websites.[489] This system allows other intermediaries to identify the same offensive content on their platforms and remove it even if the user who shared it did not use the share button.[490] Intermediaries already share digital fingerprints that allow their counterparts to identify replications of most offensive content and remove content in related contexts such as child pornography and incitement to terrorism.[491] These detection tools, however, are currently flawed and cannot properly interpret the context of expressions. Inaccurate interpretation of context can result in over-removal that would chill legitimate content.[492] Therefore, the law should not obligate intermediaries to use these tools, and it should be an intermediary's choice to use them voluntarily. "[R]emoval of all replications of text-based [posts or comments] should not be used to automatically prevent the upload[ing] of content" and should be used narrowly only for absolute falsehoods and identical replications.[493] It should only be used for detection once the offensive content has already been

---

488. *See* Chesney & Citron, *supra* note 35, at 1799 (" 'ISPs and social networks with millions of postings a day cannot plausibly respond to complaints of abuse immediately, let alone within a day or two,' yet 'they may be able to deploy technologies to detect content previously deemed unlawful.' ").

489. Lavi, *supra* note 32, at 2673.

490. For further discussion on such technology and its uses, see Rafal-Kuchta, *The Hash—A Computer File's Digital Fingerprint*, NEWTECH.LAW (Oct. 9, 2017), https://newtech.law/en/the-hash-a-computer-files-digital-fingerprint; Susan Klein & Crystal Flinn, *Social Media Compliance Programs and the War Against Terrorism,* 8 HARV. NAT'L SEC. J. 53, 79–81 (2017); Danielle Keats Citron, *Sexual Privacy,* 128 YALE L.J. 1870, 1955–58 (2019); and Klonick, *supra* note 297, at 2429–30 ("Though there have recently been steps forward in using artificial intelligence to screen for things like extremism and hate speech, the vast majority of this system works by matching the uploaded content against a database of already known illegal or impermissible content using what is known as 'hash technology.' "). It should be noted that this technology is less efficient in removing replications of text-based content, which require more sensitivity to context, rather than pictures.

491. *See generally* Fioretti, *supra* note 443.

492. *See* Citron, *supra* note 282, at 1057–58 (referring to the possible chilling effect of digital fingerprints, especially if intermediaries adopt it in the shadow of law in order to avoid regulation); KELLER, *supra* note 235, at 23–26 (noting that the filters might not be sufficiently context-sensitive in determining whether content is legal).

493. Lavi, *supra* note 32, at 2674; *see also, e.g.*, Case C-18/18, Glawischnig-Piesczek v. Facebook Ir. Ltd., ECLI:EU:C:2019:821, ¶37 (Oct. 3, 2019) (discussing removal of content previously declared unlawful). The ambiguity of equivalence can lead to removal of legitimate content such as parody and satire. To prevent this consequence, intermediaries should be encouraged to use technological tools only to remove identical content to unlawful content.

published on the platform.[494] Furthermore, human oversight should be incorporated to prevent platforms from removing legitimate content.[495]

Second, it can be argued that imposing an obligation on intermediaries to embed efficient removal in the design of the platform's code may lead to a more extensive chilling effect when this obligation is combined with a regular "notice-and-takedown" regime or an immunity regime. This solution may lead to the removal of legitimate comments on the original post as a by-product of its removal. Arguably, this solution disrupts the balance between protection of reputation and freedom of expression and results in a *disproportionate* chilling effect. Indeed, the proposed solution can exacerbate the removal of legitimate content. A closer look, however, reveals that this balance is still maintained because dissemination of offensive content also exacerbates harm to reputation. Moreover, removal occurs *ex post facto*, meaning that the content receives exposure until the victim complains. Removal may be requested only after a while or may not be requested at all. Thus, the potential chilling effect remains proportionate.

Third, it can be argued that imposing legal obligations on intermediaries to embed efficient removal mechanisms to their code might result in a chilling effect on developing advanced technological features for sharing content and other types of innovation, which is an undesirable outcome.[496] This outcome is unlikely, however, because social network intermediaries profit from features that facilitate content sharing. These mechanisms allow interpersonal communication and enhance participation and advertising profits. Thus, a cost-benefit analysis would incentivize intermediaries to continue developing innovative content-sharing features alongside implementation of the proposed code-based solution to attract more users and garner more profits. Concerns that incentives to implement code-based solutions might hinder future innovation are also unwarranted. The proposed solution does not obligate intermediaries that embed content sharing features to prefer

---

494. Lavi, *supra* note 32, at 2674.

495. *Id.*; *see also* GILLESPIE, *supra* note 297, at 98–100; CTR. FOR DEMOCRACY & TECH., MIXED MESSAGES? THE LIMITS OF AUTOMATED SOCIAL MEDIA CONTENT ANALYSIS 4 (2017) ("Today's tools for automating social media content analysis have limited ability to parse the nuanced meaning of human communication, or to detect the intent or motivation of the speaker. . ."); FILIPPO RASO, HANNAH HILLIGOSS, VIVEK KRISHNAMURTHY, CHRISTOPHER BAVITZ & KIMBERLY LEVIN, BERKMAN KLEIN CTR., ARTIFICIAL INTELLIGENCE & HUMAN RIGHTS: OPPORTUNITIES & RISKS (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3259344.

496. Scholarly work indicates a linkage between lenient liability regimes and innovation. *See* ANUPAN CHANDER, THE SILVER SILK ROAD: HOW THE WEB BINDS THE WORLD TOGETHER IN COMMERCE 57 (2013); Chander, *supra* note 186, at 667–69.

one technology over another. The focus of the conditioned safe haven is on efficient removal. This solution allows flexibility with respect to the technology at use. Thus, the risk of hindering innovation is relatively low.

Fourth, it can be argued that imposing an obligation to encode efficient removal mechanisms can be abused to undermine social campaigns. This regime risks removal of all content that elicits complaints, even if it is not false. "The chilling effect of this system [might result in] a veto power granted to anyone who has an interest [in] silenc[ing] speech, including legitimate criticism."[497] Removing all replications along with the original post is likely to make it more difficult to reveal offensive behavior that infringes on individual rights and the public interest, such as sexual harassment or corruption.[498] Even without amending § 230 and outlining a "notice-and-takedown" regime, filing an action against the original publisher or warning that failure to remove the post would lead to a lawsuit, might lead to removal of the post. Embedding efficient removal mechanisms in the code would censor all replications of the post that were shared via the share button even if the post and the replications were legitimate, for example, posts and replications that benefit from legal defenses to defamation. This may infringe on freedom of speech and the search for truth.

Efficient removal mechanisms can be abused to silence social campaigns and movements that promote the public interest, however, even social campaigns can at times be abused to spread false allegations and rumors when they go viral.[499] Due to the immense harm caused by viral dissemination of false allegations and rumors, efficient removal mechanisms are of particular importance. Indeed, removing all replications of a shared post may exacerbate the chilling effect when legitimate posts are removed. Yet, viral falsehoods cause tremendous harm to a victim's reputation and infringe on the public interest in reliable information and other *honest* social campaigns. Efficient removal mechanisms preserve the

---

497. Lavi, *supra* note 54, at 887.

498. *See, e.g.*, Jamillah Bowman Williams, Lisa Singh & Naomi Mezey, *#MeToo as Catalyst: A Glimpse into 21st Century Activism*, 2019 U. CHI. LEGAL F. 371, 371 (2019) (discussing how the #MeToo movement has used Twitter to spread sexual harassment and abuse allegations).

499. *See* Anita Raj, *Worried About Sexual Harassment – or False Allegations? Our Team Asked Americans About Their Experiences and Beliefs*, THE CONVERSATION (May 13, 2019), https://theconversation.com/worried-about-sexual-harassment-or-false-allegations-our-team-asked-americans-about-their-experiences-and-beliefs%E2%80%A6 ("[F]alse allegations of sexual harassment and assault against high-profile individuals are a growing public concern. . . . [O]ne in 20 women and one in 12 men felt that most or all of the allegations in recent high-profile cases were 'false and that accusers are purposefully lying for attention or money.'").

balance between freedom of expression, protection of reputation, and public interest in the viral age, while avoiding a disproportionate chilling effect.

Moreover, removal from all profiles and locations is *ex#post facto*. Thus, the post and its replications would remain available to the public at least until someone requests that the intermediary or publisher take them down. The gravity of harm that can be caused by viral dissemination of false allegations justifies efficient removal mechanisms, even if legitimate posts might be mistakenly removed *ex post facto*.

Fifth, it can be argued that efficient removal mechanisms could be abused by political or ideological groups that use these mechanisms to silence competing political campaigns, thereby causing injury to the marketplace of ideas.[500] Although efficient removal mechanisms can be abused by political and ideological groups to silence opposing ideologies, adopting this solution leads to a better balance between rights than its absence, where falsehoods disseminate virally without redress. In addition, the result is not as severe as might appear at first glance, since legal liability would not be imposed on the intermediary (even if legislators amend § 230) or the publisher for avoiding the removal of *legitimate* posts upon notice. Publishers and intermediaries might risk liability and avoid removing posts that are not absolutely false since they are protected by legal defenses to defamation. They might risk liability in order to promote their ideology or enhance profits.[501] In fact, not every item that users complain about is removed, as media giants use moderators that review complaints about offensive content and do not automatically remove every post upon complaint.[502]

F. *A Remark on Smartphone Social Network Applications*

This Article focuses on online intermediaries. Therefore, social network applications for smartphones, such as WhatsApp, are beyond its scope. Yet, dissemination of falsehood through smartphone apps is very common. A "notice-and-takedown" regime

---

500. In a related context, people abuse reporting systems on offensive contact to silence legitimate political campaigns. *See* VAIDHYANATHAN, *supra* note 49, at 37–39.

501. In the context of the EU "right to be forgotten" and removal of links to irrelevant search results on EU citizens, "Google received over 160,000 removal requests and denied a majority (approximately 58%) of them." Edward Lee, *Recognizing Rights in Real Time: The Role of Google in the EU Right to Be Forgotten*, 49 U.C. DAVIS L. REV. 1017, 1024 (2016).

502. Klonick, *supra* note 298, at 2433 ("[M]oderators are trained to determine if content violates the Community Standards. If a moderator determines that reported content is in violation, it is removed from the site; all content not found in violation remains published on the platform."); *see also* GILLESPIE, *supra* note 297, at 120–28.

cannot apply to them, though. Apps are similar to common carriers; their programmers are not normally responsible for defamatory statements transmitted through them since they have no editorial control.[503]

Nevertheless, the proposed solutions, which focus on the design stage, are not limited to online intermediaries and can be applied to related smartphone applications. Thus, application providers can program the application to nudge users for accountability as well. Similar to intermediaries, many application providers can be incentivized to adopt this solution as part of their business model because offensive content is a potential threat to their profits.

The second solution of efficient removal is expected to require a significant change. Unlike Facebook and Twitter, the popular messaging app WhatsApp, which boasts hundreds of millions of active users,[504] was not designed to allow efficient removal. It was not until 2017 that WhatsApp started to allow users to delete messages from recipient's devices. At first, the framework for deletion was seven minutes from message transmission, and it was gradually extended to thirteen hours.[505] If the recipient shares the content within this timeframe, the message is not likely to be deleted from the devices through which the message was shared and can still spread widely. Moreover, the removal mechanism does not currently delete all types of media from recipient devices.[506] Technology should not be taken for granted; therefore, the proposed solution of efficient removal should be developed for smartphone apps as well.

Since messages in apps are end-to-end encrypted, apps neither operate editorial control over nor bear responsibility for the content.[507] In this context, the victim of transmitted defamatory remarks would have to request removal from the original publisher. Designing the sharing feature of apps in a way that allows removal

---

503. *See* Perry & Zarsky, *supra* note 222, at 7. For further discussion of common carrier status, see Lavi, *supra* note 54, at 865.

504. *Number of Monthly Active Whatsapp Users Worldwide from April 2013 to February 2016 (in Millions)*, STATISTA (Apr. 30, 2020), http://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/; *Number of Daily Active WhatsApp Status Users from 1st Quarter 2017 to 1st Quarter 2019* (*in Millions*), STATISTA (Jan. 8, 2020), https://www.statista.com/statistics/730306/whatsapp-status-dau/ (referring to 500 million daily active users).

505. *See* Manoj Sharma, *WhatsApp Sets Over 13-hour Window to Delete Message for Everyone Permanently*, BUS. TODAY (Oct. 15, 2018), https://www.businesstoday.in/technology/news/whatsapp-sets-13hour-window-to-delete-message-for-everyone-permanently/story/285164.html.

506. *See* Mohit Kumar, *WhatsApp 'Delete for Everyone' Doesn't Delete Media Files Sent to iPhone Users* (Sept. 16, 2019), https://thehackernews.com/2019/09/whatsapp-delete-for-everyone-privacy.html.

507. *See* Citron, *supra* note 115, at 1088–91 (criticizing this result and arguing that design choices that amounted to a failure to take reasonable steps to prevent or address unlawful uses of services should not enjoy § 230 immunity).

from all the devices when the original publisher deletes the message, provides the publisher with an efficient way to remove his statement. The publisher can remove the defamatory statement in response to the victim's warning before a legal defamation suit, or upon his own regret.[508] Furthermore, since the publisher uses his mobile phone number, the victim can efficiently identify him and file a libel suit directly against him. In cases where the original publisher refuses to remove the statement, the plaintiff can seek a court injunction against him for removal of the message.

This technological solution and similar ones have already been adopted by some app providers. Messages sent through the popular app Snapchat are deleted by default after a few seconds from both the sender's and the recipient's devices.[509] While, there are ways to save posts on the app, for example, by taking screenshots or using other applications to undo the protection,[510] it is likely that most users share information by using the sharing feature embedded in the app and do not bypass it because users tend to adhere to technological defaults.[511] Therefore, most messages are likely deleted with the removal of the original post.

Similar to Snapchat, different technological supplements to a user's e-mail can "un-send" e-mails.[512] The emergence of these apps that limit dissemination might indicate a shift in market preferences to protect reputations and the general public interest. Arguably, other driving forces outside of the law governing dissemination and formal regulation would be redundant.[513] But I believe that there is reason to doubt that this would be the outcome in all cases. Similarly to intermediaries, there are different kinds of apps with different business models and attitudes towards content, and not all of them allow for efficient removal. There may be a need to incentivize app providers to adopt this solution. Providers of apps that are end-to-end encrypted lack editorial control over state-

---

508. The original disseminator may agree to remove a statement that has gone out of control if he did not spread it with malice but rather believed it to be true and discovered that it was inaccurate later on. Alternatively, the original disseminator may remove the statement in response to a libel suit.

509. *See* Tal Zarsky, *The Privacy–Innovation Conundrum*, 19 LEWIS & CLARK L. REV., 115, 167 (2015).

510. *See* HOOFNAGLE, *supra* note 354, at 135 (explaining that the FTC charged Snapchat with deceiving its users for claiming that messages sent to others would be automatically deleted even though popular applications were available to undo the protection); *In re* Snapchat Inc., FTC File No C-4501 (Dec. 23, 2014).

511. *See* THALER & SUNSTEIN, *supra* note 30, at 8 (explaining that "people have a strong tendency to go along with the status quo or default option").

512. *See, e.g.*, CRIPTEXT, https://www.criptext.com [https://perma.cc/N9UT-WJGG] (last visited Dec. 20, 2020); VIRTRU, https://www.virtru.com/gmail-encryption/recall-gmail-messages/ [https://perma.cc/YM7L-M4W9] (last visited Dec. 20, 2020).

513. *See* Zarsky, *supra* note 509, at 167.

ments transmitted over them. Thus, a conditioned "notice-and-takedown" regime would not be applicable. Mandatory design standards for apps are a possible solution, but this may not be an optimal solution with respect to apps. The range of uses of social network apps may be different from online platforms. Thus, there is no need for efficient content removal mechanisms with respect to *all* types of social network apps.[514] Design standards may not be sensitive to the differences between apps and may lack flexibility, hindering efficiency and chilling innovation.

Furthermore, end-to-end encrypted apps for secured messaging that are used by mobile devices, such as WhatsApp are considered closed spaces relative to online platforms. The content transmitted through them is less searchable, and, by default, it is difficult to find it by search engines.[515] Although falsehoods can spread among WhatsApp groups rapidly, the scope of harm may be narrower and the necessity of mandates with respect to apps is unclear. Any decision about legal regulation of apps would require evaluation of harm, the likelihood for private ordering, and the potential influences of regulation on innovation. For the time being, this Article leaves the question of mandatory incentives for app providers open.

CONCLUSION

Dissemination of content is more common today than ever before. The rise of social network platforms and their new information sharing features have increased the circulation of content exponentially by making sharing as easy as the click of a button. Each and every individual can spread almost any message he wants, as long as he could get a crowd to listen. Within seconds, a message or a post can travel around the world and be viewed by thousands of users. While online dissemination affords many benefits, it can infringe upon important values. Content disseminated may include falsehoods, defamation, and risk: harm to the victim's reputation and his standing as an equal member of society, economic loss,

---

514. For example, app providers may develop social network apps for professional uses in specific organizations. *See, e.g.*, Chris Welch, *Microsoft's Latest 'Garage' Project Is a Dead-Simple Email App for iPhone*, THE VERGE (July 22, 2015), https://www.theverge.com/2015/7/22/9013885/microsoft-send-email-app-announced.

515. *See* Lisa Vaas, *Google Stops Indexing WhatsApp Chats; Other Search Engines Still at It*, NAKED SECURITY (Feb. 25, 2020), https://nakedsecurity.sophos.com/2020/02/25/google-stops-indexing-whatsapp-chats-other-search-engines-still-at-it/ (noting that secured private messaging should not be found on search engines, whereas messages sent via group chats should be difficult to find).

and severe emotional harm. Moreover, the damage caused by dissemination of falsehoods extends beyond private individuals, as it becomes more difficult to separate truth from falsehood and engage in truthful discussions on matters of public importance, thereby impinging upon the general public interest.

The widespread dissemination of falsehoods online poses complex challenges to legislators, courts and policy makers. This Article has endeavored to identify potential remedies to meet the challenges posed by spreading falsehoods on social networks. Policy discussions on this issue thus far have failed to account for human motivations to spread falsehoods, the social dynamics of networks, and the influence of intermediaries on the flow of information. Only by understanding the social dynamics and motivations behind spreading falsehoods can a solution be developed. Analysis of the way falsehoods spread constitutes an indispensable step towards fully acknowledging the challenges and potential solutions.

As the influence of intermediaries on the flow of information becomes a fundamental aspect of any platform, their duties should be reconsidered. The proposed solutions for mitigating the harm of dissemination of falsehoods focus on the design stage of platforms. The first solution utilizes choice architecture and nudges to dissuade users from sharing falsehoods *ex ante*. The second solution utilizes code to allow efficient removal of falsehoods from every profile and location where they were shared *ex post facto*. Under this solution, unless intermediaries develop efficient removal techniques for mitigating harmful content shared via publish, share, and re-tweet buttons, they would lose § 230 immunity.

Outlining fair and efficient regulation of dissemination of falsehoods is one of the most prominent challenges of the digital era. The analysis and solutions proposed here have vast potential to meet this challenge and improve the current regulatory regime. The solutions are important, especially in light of the influence of falsehoods on political views, voters election results, and democracy. This Article is not the last word on the topic, as digital dissemination creates new risks to reputation and the public interest. There are further avenues of analytic inquiry on the power of intermediaries and their influence on the flow of information. Should intermediaries bear additional duties and obligations in light of their influence? And if so, what should be the normative framework and scope of these duties? What should be the scope of § 230 CDA's immunity? Should the law impose transparency and due process obligations on intermediaries, even though they are private actors? These challenges and others await another day.