Harvard Journal of Law & Technology
Volume 34, Number 1 Fall 2020

ENJOINING NON-LIABLE PLATFORMS

*Maayan Perel**

TABLE OF CONTENTS

---

* Assistant Professor at Netanya Academic College, Faculty of Law, and a Senior Research Fellow at the Haifa Center for Law & Technology, University of Haifa, Faculty of Law. This Research was supported by the Israel Science Foundation (grant No. 1820/17).

I. INTRODUCTION

The proliferation of online content involves dozens of platforms carrying material to countless recipients.[1] Platforms' role in spreading content raises serious concerns regarding their responsibility to restrict harmful speech.[2] Governments, civil societies, and activists around the globe contend that platforms should do more to protect our online sphere from poisonous content, such as hate speech and copyright infringement.[3] Questions of platforms' responsibility and liability are at the center of this discourse. Specifically, the safe harbor accorded to platforms under § 230 of the Communications Decency Act ("CDA") is under fire.[4]

In the United States, platforms enjoy strong and rather stable immunities from acts of infringement caused by their users. The CDA exempts Internet Service Providers ("ISPs") and some other online intermediaries from certain kinds of third party liability by

---

1. MARTIN HUSOVEC, INJUNCTIONS AGAINST INTERMEDIARIES IN THE EUROPEAN UNION, 9 (2017).

2. *Id.*

3. *See, e.g.*, *Facebook Must Delete Hate Postings, Austria Court Rules*, BBC NEWS (May 9, 2017), https://www.bbc.com/news/world-europe-39852623 [https://perma.cc/4AF5-483Q]; Davey Alba, *A Court Order to Terminate Hate Speech Tests Facebook*, WIRED (Sep. 5, 2017), https://www.wired.com/2017/05/court-order-terminate-hate-speech-tests-facebook [https://perma.cc/J8SG-J5NP]; David Meyer, *Facebook Can Block Hate Speech, Even if It's Not Illegal, Court Rules*, ZDNET (Sep. 18, 2018), https://www.zdnet.com/article/facebook-can-block-hate-speech-even-if-its-not-illegal-court-rules/ [https://perma.cc/845G-TS6F]; Matt Reynolds, *What Is Article 13? The EU's Divisive New Copyright Plan Explained*, WIRED (May 24, 2019), https://www.wired.co.uk/article/what-is-article-13-article-11-european-directive-on-copyright-explained-meme-ban [https://perma.cc/3F55-8U5M].

4. *See, e.g.*, Matt Laslo, *The Fight Over Section 230 — and the Internet as We Know It*, WIRED (Aug. 13, 2019), https://www.wired.com/story/fight-over-section-230-internet-as-we-know-it/ [https://perma.cc/ZE8X-STDL]; Taylor Hatmaker, *Nancy Pelosi Warns Tech Companies That Section 230 Is 'in Jeopardy,'* TECHCRUNCH (Apr. 12, 2019, 3:35 PM), https://techcrunch.com/2019/04/12/nancy-pelosi-section-230/ [https://perma.cc/H9AH-NQJU]; Emily McPhie, *Part II: Senators Josh Hawley and Ted Cruz Want to Repeal Section 230 and Break the Internet*, BELTWAY BREAKFAST (Aug. 20, 2019), https://www.beltwaybreakfast.com/its-all-connected/2019/08/20/part-ii-senators-josh-hawley-and-ted-cruz-want-to-repeal-section-230-and-break-the-internet/ [https://perma.cc/Y6VU-DMKB]; Daisuke Wakabayashi, *Legal Shield for Websites Rattles Under Onslaught of Hate Speech*, N.Y. TIMES (Aug. 6, 2019), https://nyti.ms/2TcXY0S [https://perma.cc/RPB4-J52C]. *But see* Elliot Harmon, *Changing Section 230 Would Strengthen the Biggest Tech Companies*, N.Y. TIMES (Oct. 16, 2019), https://nyti.ms/31rtMBG [https://perma.cc/TD5X-LEY2]. *See also* 47 U.S.C. § 230.

determining that they are not "the publisher or speaker of any information provided by another information content provider."[5] Similarly, the Digital Millennium Copyright Act of 1998 ("DMCA") bars indirect copyright liability for ISPs who are acting only as a conduit and limits liability for web hosting and other service providers if they follow a prescribed notice-and-takedown procedure.[6] Similar immunities were also enacted under internet gambling and online pharmacy laws.[7]

While some legal scholars seem to be skeptical that making platforms liable for harmful activity on their services would be the right cure against poisonous content, others advocate that the interpretation of § 230 immunity is too broad, leaving "victims of online abuse with no leverage against site operators whose business models facilitate abuse."[8]

The U.S. legislature has also attempted to address the issue of harmful online content through the lens of platform liability. Congress has recently reduced § 230's immunity, and in 2018 passed the Allow States and Victims to Fight Online Sex Trafficking Act of 2017 ("FOSTA"), "designed to attack the online promotion of sex trafficking victims, in part, by, reducing § 230's scope."[9] Additionally, in October 2019 the House Subcommittee on Communications and Technology and the Subcommittee on Consumer Protection and Commerce held a hearing titled "Fostering a Healthier Internet to Protect Consumers." The main focus of the hearing was § 230 and whether it should be amended considering the scope of harmful online activity that the platforms have failed to address.[10]

---

5. 47 U.S.C. § 230(c)(1).

6. 17 U.S.C. § 512(a), (c)–(d).

7. *See, e.g.*, Unlawful Internet Gambling Enforcement Act of 2006, 31 U.S.C. §§ 5361–5367; Ryan Haight Online Pharmacy Consumer Protection Act of 2008, 21 U.S.C. §§ 829, 802.

8. *See* Eric Goldman, *Why Section 230 Is Better Than The First Amendment*, 95 NOTRE DAME L. REV. 33, 34 (2019); *see also* Niva Elkin-Koren, Yifat Nahmias, Mayaan Perel, *Is It Time to Abolish Safe Harbor? When Rhetoric Clouds Policy Goals*, 31 STAN. L. & POL'Y REV. 1, 7, 9–11 (2020). *But see* Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 404 (2017); Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986, 1002–05 (2008) (contending that § 230 fails to take into account circumstances in which the relative interests and incentives of speakers and intermediaries justify imposing responsibility on the intermediary); Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293, 330–31, 349 (2011) (arguing that intermediaries should not receive immunity when they are acting with the incentives of an original speaker or when the form of liability at issue is uniquely applicable to intermediaries rather than original speakers).

9. Eric Goldman, *The Complicated Story of FOSTA and Section 230*, 17 FIRST AMEND. L. REV. 279, 280 (2019).

10. *Congress Holds a Hearing on Section 230 of the Communications Decency Act*, NEWS MEDIA ALLIANCE (Oct. 17, 2019), https://www.newsmediaalliance.org/congress-

Noticeably missing from the current discourse over harmful online content in the U.S. is the possibility of harnessing platforms' potential *enforcement* capabilities, regardless of questions of liability, to reduce some of the harms caused by online speech. This is mainly because in the U.S., the Free Speech Clause of the First Amendment restricts government regulation of private speech.[11] Despite concerns that "the real threat to free speech today comes from private entities such as Internet service providers, not from the Government," interfering with the editorial discretion of platforms is seen as a violation of platforms' First Amendment rights.[12]

However, platforms are a natural point of controlling the substance of online communications and hence are capable of preventing the dissemination of unlawful content.[13] They have the means "to intervene in the circulation of abhorrent content and at the moment of abhorrent behavior."[14] In Europe, the engagement of online intermediaries in enforcing the rights of individuals allegedly harmed by online speech has recently shifted to what Martin Husovec names "accountability without liability."[15] That is, non-liable online platforms in the European Union ("EU") are increasingly forced to assist rightsholders and target speech or speakers that violate their rights, even though these platforms were not involved in any unlawful way in disseminating that speech.[16]

In the U.S., however, the role of platforms in addressing illegal content is still defined according to liability theories. Content removals by non-liable platforms are currently conducted mostly on a voluntary basis.[17] Governmental agents and other authorized reporters can file requests with various platforms to remove allegedly illicit

---

holds-a-hearing-on-section-230-of-the-communications-decency-act/ [https://perma.cc/V9LF-D94W].

11. Tim Wu, *Is the First Amendment Obsolete?*, 117 MICH. L. REV. 547, 568–69 (2018); *see also* Manhattan Cmty. Access Corp. v. Halleck, 139 S. Ct. 1921, 1926 (2019) (holding that the Free Speech Clause of the First Amendment of the US Constitution prohibits only governmental, not private, abridgment of speech).

12. U.S. Telecom Ass'n v. FCC, 855 F.3d 381, 434 (D.C. Cir. 2017) (Kavanaugh, J., dissenting); *see also* DAPHNE KELLER, WHO DO YOU SUE? STATE AND PLATFORM HYBRID POWER OVER ONLINE SPEECH 2, 4 (Hoover Inst., Aegis Series Paper No. 1902, 2019), https://www.hoover.org/sites/default/files/research/docs/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech_0.pdf [https://perma.cc/YH3Z-3ELG].

13. *See* Jonathan Zittrain, *Internet Points of Control*, 44 B.C. L. REV. 653, 655 (2003).

14. Tarleton Gillespie, *Regulation of and by Platforms*, *in* SAGE HANDBOOK OF SOCIAL MEDIA, 23 (Jean Burgess, Thomas Poell, and Alice Marwick eds., 2017).

15. HUSOVEC, *supra* note 1, at 10.

16. *Id.* at 9.

17. *But see Removal Requests*, TWITTER TRANSPARENCY REPORT, https://transparency.twitter.com/en/removal-requests.html#removal-requests-jan-jun-2019 [https://perma.cc/5BR8-EHYD] (stating that under Twitter's Country Withheld Content Policy, Twitter is obliged to remove content for legal reasons).

content from their services.[18] Nevertheless, platforms are not legally bound by such removal requests, so they can elect to partially or completely decline them.[19] Furthermore, platforms also engage in voluntary content moderation. They may enable or disable access to content by removing or blocking controversial content, or by terminating the accounts of particular speakers.[20] In this respect, they follow their internal policies regarding objectionable content (e.g., community guidelines) to satisfy their users and assure they spend as much time as possible on their services.[21]

Unless directly, vicariously, or contributorily liable for the harms caused by illicit content, platforms cannot be forced by courts to actively remove it.[22] This current state of affairs, however, completely ignores the natural position of platforms as doormen who govern the free flow of online information.[23] It fails to harness the tremendous power platforms could exercise as authorized law enforcers. As this paper sets forth, this failure is rooted in two main legal barriers. The first is procedural and concerns the statutory restriction on enjoining non-liable third parties. For decades, bedrock rules of equity and due process have defended non-liable third parties from being enjoined by courts since they are "strangers to the litigation."[24] In *Blockowicz v. Williams*,[25] the Seventh Circuit ruled that the fact that platforms are "technologically capable of removing" questionable content "does not render [their] failure to do so aiding and abetting," which is what is required under Rule 65 of the Federal Rules of Civil Procedure in order to enjoin non-parties.[26] Nevertheless, it is hard to ignore the fact

---

18. *See*, *e.g.*, *Government Requests to Remove Content*, GOOGLE TRANSPARENCY REPORT, https://transparencyreport.google.com/government-removals/overview [https://perma.cc/AF7F-364A]; *Removal Requests*, TWITTER TRANSPARENCY REPORT, https://transparency.twitter.com/en/removal-requests.html [https://perma.cc/74UK-ZK6T]; Transparency Report 2018, REDDIT, https://www.redditinc.com/policies/transparency-report-2018 [https://perma.cc/4L9L-BKMW].

19. According to Google's Transparency Report, 60% of the requests from governmental agencies or law enforcement agents were partially or completely acted upon. Government requests to remove content, *Removal Requests by Country/Region*, GOOGLE TRANSPARENCY REPORT, *supra* note 18, (search for "United States" in the "Removal Requests by County/Region" subsection).

20. *See* KELLER, *supra* note 12, at 2.

21. Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUM. L. REV. 2011, 2022 (2018); Danielle Keats Citron & Helen Norton, *Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age*, 91 B.U. L. REV. 1435, 1454–55 (2011).

22. *See* David S. Ardia, *Freedom of Speech, Defamation, and Injunctions*, 55 WM. & MARY L. REV. 1, 17 n.64 (2013) (citing Bobolas v. Does, No. CV-10-2056-PHX-DGC, 2010 WL 3923880, at *2 (D. Ariz. Oct. 1, 2010) (refusing to enjoin GoDaddy.com because it was not an agent of the defendant)).

23. *See* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1603–04 (2018).

24. Bobolas v. Does, No. 1:10-CV-2056, 2010 WL 3923880, at *2 (D. Ariz. Oct. 1, 2010); *see infra* Section IV.A.1.

25. 630 F.3d 563 (7th Cir. 2010).

26. *Id.* at 568; FED. R. CIV. P. 65.

that often times platforms, albeit not liable, are effectively the only entities in the position to stop the accelerating harm caused by illegitimate content going viral.[27]

The second legal barrier to making speech regulation by platforms mandatory is the absence of a formal legal procedure that would allow law enforcement agents to seek court orders that would force platforms to remove illegal content.[28] Outside the area of speech regulation, legal procedures that demand action on the part of platforms do exist. The Stored Communication Act ("SCA"), which established a legal procedure for a governmental entity seeking action on the part of the platform, is one example.[29] Nevertheless, within the area of speech regulation, enforcement by non-liable platforms is largely based on out-of-court submissions made directly to the platforms and is therefore mostly voluntary.[30]

This paper advocates making speech regulation by platforms mandatory. It promotes scrutinized removal of illegal content by non-liable platforms, governed by ongoing judicial review.[31] Specifically, this paper focuses on platforms' *ability* to remove illegal content, rather than on platforms' *liability* for the proliferation of such content. Accordingly, the paper proposes two legal fixes: first, allowing civil injunctions against non-liable platforms that enable the dissemination of tortious content, and second, establishing an open and transparent statutory procedure that will allow designated law enforcement agents to request courts to order platforms to remove content that was proved to be illegal by clear and convincing evidence.

The discussion proceeds as follows: Part II presents the governing theory of platform liability which is meant both to prevent and to address actions (or inactions) of platforms that unlawfully contribute to the dissemination of tortious content. Given the broad immunities

---

27. *See* Courtney Brown, *Caught in a Bind: Reassuring Judicial Authority to Bind Non-Party Search Engines under Rule 65 in Counterfeit Goods Cases*, 32 CARDOZO ARTS & ENT. L.J. 257, 259 (2013) ("Because courts have not been able to locate or seize counterfeiters' assets, judges . . . have been ordering that . . . search engines must block access to the counterfeiting sites and . . . sites registered by recurring counterfeiters in the future by excluding those sites from their search results."); *see also infra* Section II.B.1.

28. *See* Jacquelyn E. Fradette, *Online Terms of Service: A Shield for First Amendment Scrutiny of Government Action*, 89 NOTRE DAME L. REV. 947, 980–81 (comparing the statutory procedures for government information access requests with the absence of similar procedures for takedown requests).

29. *See* 18 U.S.C. §§ 2701–2711 (2006).

30. *See, e.g.*, Fradette, *supra* note 28, at 967 (providing an example of the voluntary nature of government takedown requests to Google).

31. This is a crucial way in which this paper's proposal deviates from recent European initiatives. Indeed, "the European Commission aligns its strategy for online platforms to a globalized, ongoing move towards privatization of enforcement online through algorithmic tools. This process may advance an amorphous notion of responsibility that incentivizes intermediaries' self-intervention to police allegedly infringing activities in the internet." Giancarlo F. Frosio, *Reforming Intermediary Liability in the Platform Economy: A European Digital Single Market Strategy*, 112 NW. U. L. REV. 19, 19–20 (2017).

accorded to platforms under § 230 of the CDA, this Part further explains the shortcomings of platform liability in addressing harmful online content. Part III introduces the current enforcement-based regime for speech regulation by platforms. It discusses content removals that are based on governmental removal requests submitted directly to platforms or on content moderation practices and highlights their voluntary nature. Platforms are not obliged to engage in these regulatory efforts, and therefore it is impossible to rely on them to sufficiently and legitimately address illegal content. This Part concludes that the removal of illicit content by platforms should become mandatory, open, and subject to judicial review.

Next, Part IV discusses the first recommended legal fix that is needed to achieve this goal: allowing courts to issue civil injunctions against non-liable platforms, ordering them to remove tortious content. This Part explains how the adoption of this fix may be obstructed by different legal barriers, including procedural due process, the doctrine of prior restraint, and the platforms' legitimate business interests. Subsequently, this Part shows that by crafting a constrained legal procedure for issuing such injunctions, the barriers discussed can be successfully overcome. Particularly, it is necessary to provide adequate notice to affected platforms as well as an opportunity to object to the injunction. It is also important to limit the applicability of the procedure to cases where there is no other way to remove the tortious content but to harness the platform's technological capabilities.

Part V then focuses on the second recommended fix that is needed in order to assure the removal of illegal content: enacting a new, open, and transparent statutory procedure that will allow designated law enforcement agents to file removal requests with courts in relation to content that is proved to be illegal by clear and convincing evidence. This Part discusses the constitutional barriers that could arguably interfere with this fix, including the First Amendment, the doctrine of prior restraint, and the concept of regulatory takings under the Fifth Amendment. Afterwards, this Part moves to argue that by creating a restricted legal process for removal that can only be exploited after the illegality of the content is proved by clear and convincing evidence, and by compensating platforms for their compliance costs, it is possible to remove these constitutional barriers.

Finally, Part VI provides a concluding tradeoff map that balances between the values democracies can promote by adopting the proposed fixes, including securing the rule of law, promoting the oversight of speech regulation by platforms, and protecting public safety, against the values these fixes may put at risk unless carefully and restrictedly designed, such as the free flow of information,

innovation, and flexibility. This Part further presents some safety valves that will minimize the risk to some of the values discussed in order to assure a balanced and effective regime of speech regulation. A short conclusion follows.

II. THEORETICAL FRAMEWORK: PLATFORM LIABILITY

In the United States, platforms enjoy general immunity from claims based on users' content under § 230 of the CDA.[32] Under this regime — generally considered to be one of the most important protections of free speech in the United States in the digital age[33] — Facebook, for instance, will not be held liable for a defamatory post by a user unless Facebook was directly involved in generating the defamatory content.[34]

Despite its broad interpretation, § 230 is not about absolute immunity for platforms.[35] First, it expressly excludes intellectual property law, federal criminal law, and communications privacy law from its coverage.[36] Second, while it stresses 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," it does not ban holding platforms liable as primary speakers.[37] § 230, hence, leaves some room to rely on platform liability to reduce the spread of illicit content online.[38] Nonetheless, this is still far from what is necessary to deal with illicit content online.[39] The following discussion describes the current

32. 47 U.S.C. § 230(c)(1) ("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."); Klayman v. Zuckerberg, 910 F. Supp. 2d 314, 318 (D.D.C. 2012). *But see* Fraley v. Facebook, Inc., 830 F. Supp. 2d 785, 802 (N.D. Cal. 2011) (holding that § 230 does not protect the defendant because the defendant appeared to be a content provider); Carafano v. Metro-splash.com, Inc., 207 F. Supp. 2d 1055, 1065–68 (C.D. Cal. 2002) (holding that § 230 does not protect the defendant because the defendant was a content provider in addition to being an ISP).

33. Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2313 (2014) ("Section 230 immunity . . . ha[s] been among the most important protections of free expression in the United States in the digital age.").

34. *See* David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 460 (2010).

35. *See* Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12, 19 (1st Cir. 2016) (listing federal appellate cases in which courts found causes of action against platforms by treating them as publishers or speakers of content provided by others).

36. 47 U.S.C. § 230(e).

37. 47 U.S.C. § 230(c)(1).

38. *See, e.g.*, Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1100 (9th Cir. 2009) (noting that the CDA does not declare "a general immunity from liability deriving from third-party content").

39. *See* Gillespie, *supra* note 14, at 12.

framework of liability-based speech regulation and explains its shortcomings in addressing harmful content online.

A. Direct Liability

Platforms can be held directly liable for their own illegal conduct, notwithstanding the broad immunity of § 230, as recently recognized by the Third Circuit in *Oberdorf v. Amazon.com Inc*.[40] This product liability case discussed Amazon's liability for the injuries suffered by Heather Oberdorf, a consumer who purchased a defective dog collar on Amazon.com.[41] As the court described, when Heather Oberdorf walked her dog, the ring on the dog's collar suddenly broke, causing the leash to recoil back. As a result, Heather's eyes were injured badly, and she became permanently blind in her left eye. Since the defective dog collar wasn't sold directly by Amazon.com as a vendor, the main issue was Amazon's role in effectuating the sale of products offered by third party vendors.[42]

In defending against Oberdorf's strict product liability claim, Amazon contended that "it is not a 'seller' because it merely provides an online marketplace for products sold by third-party vendors."[43] Additionally, Amazon claimed that Oberdorf's strict liability and negligence claims were barred under § 230 of the CDA.[44] The court, however, disagreed with Amazon. First, the court found that Amazon should be considered the "seller" of products offered by third party vendors, and therefore it is strictly liable for consumer injuries caused by defective goods purchased on its platform.[45] Second, the court found that claims against Amazon are only precluded under § 230 whenever "the duty that the plaintiff alleges the defendant violated derives from the defendant's status or conduct as a 'publisher or speaker.'"[46] However, Amazon is not immune against claims premised on other actions or failures in *the sales or distribution processes*.[47] Accordingly, the court concluded that "to the extent that Oberdorf's negligence and strict liability claims rely on Amazon's role as an actor in the sales process, they are not barred by the CDA."[48] However, the allegation that Amazon failed to provide or to

40. Oberdorf v. Amazon.com Inc., 930 F.3d 136, 140 (3rd Cir. 2019), *vacated on other grounds*, 818 Fed. Appx. 138.
41. *Id.*
42. *Id.*
43. *Id.* at 143.
44. *Id.* at 151–52.
45. *Id.* at 148–51.
46. *Id.* at 152 (citing Barnes v. Yahoo!, Inc, 570 F.3d 1096, 1102 (9th Cir. 2009), *as amended* (Sep. 28, 2009)).
47. *Id.* at 153.
48. *Id.*

edit adequate warnings regarding the use of the dog collar, does fall "within the publisher's editorial function," and "[f]or that reason, these failure to warn claims are barred by the CDA."[49]

Similarly, speech forums can be held directly liable for harm caused by content they have a role in publishing.[50] For instance, in the *Fair Housing Council of San Fernando Valley v. Roommates.com*[51] case, the Ninth Circuit found that Roomates.com was the creator of the content because it required subscribers to create profiles and answer questions — about themselves and preferences in roommates — regarding criteria including sex, sexual orientation, and whether they would bring children to the household.[52] Since Roomates.com became much more than a passive transmitter of information provided by others, the court held that the website could be liable for violations of the federal Fair Housing Act and California housing-discrimination laws.[53] Likewise, in *Fraley v. Facebook, Inc.*,[54] where users sued Facebook for using their profile pictures in ads, claiming a right-of-publicity violation, the District Court for the Northern District of California rejected Facebook's § 230 defense.[55] The court ruled that this was not a case where plaintiffs accused a platform for publishing tortious content, but where they accused the platform of "creating and developing commercial content that violates their statutory right of publicity."[56] Hence, the direct involvement of Facebook in the creating and posting of the infringing content makes it an information content provider, which is not protected from liability under § 230.[57]

Nevertheless, where platforms do not create or develop the infringing content themselves, they cannot be held liable for any harm it causes.[58] In fact, even when it appears that platforms are actively involved in enabling the illegal content, they are still likely to escape

---

49. *Id.*

50. *See* Harris v. Minvielle, 19 So. 925, 928 (La. 1896) (stating that "[t]alebearers are as bad as talemakers"); Dixson v. Newsweek, Inc., 562 F.2d 626, 631 (10th Cir. 1977) (stating that a "republication of false defamatory statements is as much a tort as the original publication").

51. 521 F.3d 1157 (9th Cir. 2008) (en banc).

52. *Id.* at 1165 ("[T]he party responsible for putting information online may be subject to liability, even if the information originated with a user.").

53. *Id.* at 1175.

54. 830 F. Supp. 2d 785 (N.D. Cal. 2011).

55. *Id.* at 803.

56. *Id.* at 801.

57. *Id.* at 801–02.

58. *See, e.g.*, Carafano v. Metrosplash.com, Inc., 339 F.3d 1119, 1125 (9th Cir. 2003); Doe v. MySpace, Inc., 528 F.3d 413, 418 (5th Cir. 2008); Jurin v. Google, Inc., 695 F. Supp. 2d 1117, 1122–23 (E.D. Cal. 2010); Doe IX v. MySpace, Inc., 629 F. Supp. 2d 663, 665 (E.D. Tex. 2009); Doe II v. MySpace, Inc., 96 Cal. Rptr. 3d 148, 156–59 (Cal. Ct. App. 2009).

liability. In *Dyroff v. Ultimate Software Group, Inc.*,[59] for instance, the plaintiff asserted that Ultimate Software's (now inactive) social-network website "Experience Project" used data mining and machine learning to understand "the meaning and intent behind posts" and target illegal material to individual users.[60] Specifically, Dyroff, whose 29-year-old son died from an overdose of heroin, claimed that the software applied by Ultimate Software eventually steered her son toward heroin-related discussion groups and the drug dealer who ultimately sold him a deadly fentanyl-laced heroin.[61] Although it was undisputed that the proprietary algorithms of Ultimate Software are what facilitated the connection between Dyroff's son and the dealer who sold him the fentanyl-laced heroin, the court held that Ultimate Software Group was immune under § 230 because Ultimate Software did not post the content in the dealer's ad.[62] According to the court, "providing content-neutral tools" (as opposed to discriminatory search criteria such as those used in the *Roommates* case) "to facilitate communication does not create liability."[63]

Hence, the prospect of direct liability as a means of making platforms accountable for harms that are caused by illicit content disseminated through their services are extremely narrow. Unless actively involved in creating or developing the illicit content, platforms are immune from liability under the broad safe harbor of § 230. As explained next, the doctrine of secondary liability leaves some additional room for the imposition of indirect liability on the basis of knowledge.

B. Secondary Liability and Notice-and-takedown

Platforms also face various legal claims arising from the content of third parties, including intellectual property law, antidiscrimination laws, and state tort laws. Originating in tort law, secondary liability can be imposed whenever someone who did not directly commit the legal wrong is found responsible for encouraging, facilitating, or profiting from it.[64] Courts often justify "secondary liability on economic efficiency grounds, viewing it as a means to shift injury costs to those who are in a position to prevent future injuries."[65]

---

59. No. 17-CV-05359, 2017 WL 5665670 (N.D. Cal. 2017).
60. *Id.* at *8.
61. *Id.* at *1.
62. *Id.* at *8–9.
63. *Id.* at *5.
64. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 25:23 (4th ed. 2005).
65. Mark Bartholomew & John Tehranian, *The Secret Life of Legal Doctrine: The Divergent Evolution of Secondary Liability in Trademark and Copyright Law*, 21 BERKELEY TECH. L.J. 1363, 1366 (2006).

Secondary liability comes in two forms: vicarious liability and contributory liability. Vicarious liability imposes strict liability on defendants that control or who have the right to control the direct tortfeasor.[66] Contributory liability imposes liability on distinct parties "if they acted in concert with or provided assistance or encouragement to the direct tortfeasor."[67] Furthermore, knowledge is "required for contributory liability: the contributory tortfeasor must purposefully assist the performance of a tortious act."[68]

Specifically relevant to platforms' secondary liability for harms caused by online content is intellectual property, which is expressly excluded from the coverage of § 230.[69] To overcome the knowledge hurdle under which platforms cannot be held liable for infringing content on their services unless they knew about its existence, the DMCA uses detailed procedural rules.[70] Addressing copyright only, the DMCA establishes a procedural framework to cabin culpable knowledge known as notice-and-takedown. [71] This framework essentially limits platforms' liability, provided that the platform takes down infringing content of which it has been notified and reposts content in response to claims that it is not, in fact, infringing.[72] The enactment of this intermediary safe harbor reflected a compromise between Online Service Providers ("OSPs") and copyright holders on whether the former — which are hosting and transmitting material from users "without modification" — should be treated as publishers of that material, and therefore liable for copyright infringement.[73]

OSPs opposed the DMCA safe harbor, contending that "many of the new services enabled by the Internet precluded the type of editorial involvement on which publisher liability has relied." [74] Intermediary liability was — and still is — often seen as a threat to

66. *Id*. at 1367.

67. *Id.*

68. *Id.*

69. 47 U.S.C. § 230(e)(2).

70. 17 U.S.C. § 1201; *see also* Daphne Keller, *Toward a Clearer Conversation About Platform Liability*, KNIGHT FIRST AMENDMENT INST. (Apr. 6, 2018), https://knightcolumbia.org/content/toward-clearer-conversation-about-platform-liability [https://perma.cc/N2MJ-VDK8**]**.

71. Note that in regard to other forms of harmful speech, such as defamation, the CDA immunizes service providers from liability, regardless of whether they attempt to remove potentially defamatory content. *See, e.g.*, Batzel v. Smith, 333 F.3d 1018, 1034 (9th Cir. 2003) (declaring that an operator of a listserv and website is a user of interactive computer services, entitling CDA protection from liability for publishing information provided by another information provider).

72. 17 U.S.C. § 512.

73. For a comprehensive description and analysis of the passage of the DMCA, *see* JESSICA LITMAN, DIGITAL COPYRIGHT: PROTECTING INTELLECTUAL PROPERTY ON THE INTERNET (2000).

74. JENNIFER M. URBAN, JOE KARAGANIS & BRIANNA SCHOFIELD, NOTICE AND TAKEDOWN IN EVERYDAY PRACTICE 7 (UC Berkeley Public L., Rsch. Paper No. 2755628, 2017).

free speech and the free flow of information because it pushes intermediaries to be on the safe side and remove content, even when not needed.[75] OSPs claimed that what was possible for print publishers or newsroom editors was allegedly not scalable to thousands or millions of user-generated posts, comments, or data transfers.[76] Rightsholders, on the other hand, argued that traditional publisher liability provided the right model for online intermediaries to address the vastly expanded capacity for copyright infringement on the Internet.[77]

Eventually, to compromise between these two positions, the DMCA accorded OSPs immunity from secondary liability for their users' copyright infringement, in return for OSPs' compliance with notice and takedown procedures. Under these procedures, "copyright owners can request that infringing materials be removed from online sites by sending brief 'takedown notices' to OSPs, without the expense and hassle of filing a lawsuit."[78] As to the targets of such notices, the system allows them to dispute the removal of their content by filing a counter notice and requesting that their content be reinstated.[79]

Over the past two decades, the system of notice-and-takedown has become a main mechanism for resolving copyright disputes.[80] In the years since its enactment, notice-and-takedown has turned into a robust mechanism of algorithmic copyright enforcement. As millions of takedown notices are now filed by so-called automatic "robo-notices" filers that scan the web to identify uses of copyrighted material, platforms deploy automatic machines to address them with no human intervention.[81] Increasingly, platforms adopt "DMCA Plus" measures, such as ex-ante filtering or staydown measures, to proactively prevent infringing material from making its way onto (or staying on) an OSP's system.[82]

---

75. *See* Margot E. Kaminski, *Positive Proposals for Treatment of Online Intermediaries*, 28 AM. UNIV. INT'L L. REV. 203, 205 (2012).

76. *See* Jennifer M. Urban, Joe Karaganis & Brianna Schofield, *Notice and Takedown: Online Service Provider and Rightsholder Accounts of Everyday Practice*, 64 J. COPYRIGHT SOC'Y U.S.A. 382 (2017).

77. *Id*.

78. *Id*. at 373.

79. *See* 17 U.S.C. § 512(g) (2012); *see also* § 512(f) (2012) (providing liability for damages, including costs and attorney fees, incurred as a result of the service provider relying upon such misrepresentation in removing or disabling access to material).

80. Sharon Bar-Ziv & Niva Elkin-Koren, *Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown*, 50 CONN. L. REV. 339, 342 (2018).

81. URBAN, KARAGANIS & SCHOFIELD, *supra* note 74, at 31–32; Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473, 477 (2016).

82. This can include metadata monitoring of content, which is essentially designed to search the file's metadata or other textual tags attached to it to match it to an existing

1. The Limitations of Notice and Takedown

Notice and takedown is often celebrated for affording rightsholders an efficient mechanism for removing material violating their exclusive rights in the digital world.[83] Moreover, it is seen as an important mechanism that helps "states persuade owners of private infrastructure to work with them and for them."[84] Nevertheless, it is insufficient to fully and properly address online copyright infringement. To begin with, the system of notice-and-takedown does not mandate the removal of allegedly infringing content, but rather merely encourages such removals. Indeed, OSPs are free to decline to comply with a takedown notice, consequently losing the protection of the statutory safe harbor.[85] As a system operating outside the courthouse, it is largely based on platforms' internal risk assessment, and not on mandatory judicial enforcement.[86] Only if a rightsholder whose notice was ignored elects to proceed with her claim and file a suit in court, and the court then finds that the designated content is indeed infringing, might the platform be ordered to remove the content. Since the cooperation of platforms in removing infringing content is the "heart and bones" of online copyright enforcement, it only makes sense to assure platforms' enforcement potential is exploited to its maximum.

Additionally, under the DMCA, platforms have to "expeditiously" remove the allegedly infringing content; however, it is not clear how responsive they are actually required to be to comply with this threshold.[87] Anecdotal evidence indicates that it might take platforms several weeks — or months — to process and respond to a takedown notice.[88] This should not come as a surprise considering the

---

catalog of files. *See, e.g.*, Melanie Ehrenkranz, *The Best NSFW Instagram Hashtags Use Special Characters to Hide Porn*, MIC (July 14, 2016), https://mic.com/articles/148675/the-best-nsfw-instagram-hashtags-use-special-characters-to-hide-porn-enjoy#.1Drt7g5Ua [https://perma.cc/J2X3-UHHV]; *see also* Urban, Karaganis & Shofield *supra* note 76, at 383.

83. *See* URBAN, KARAGANIS & SCHOFIELD, *supra* note 74, at 56.

84. *See* Balkin, *supra* note 33, at 2311.

85. *Id.* at 2314 ("Section 512(g) of the DMCA offers companies that host content a safe harbor only if they agree to a notice-and-takedown scheme.").

86. Anecdotal evidence proves that OPSs do ignore some takedown notices in practice. *See, e.g.*, URBAN, KARAGANIS & SCHOFIELD, *supra* note 74; Nate Hoffelder, *Internet Archive Ignores DMCA Notices*, DIGITAL READER (Feb. 22, 2018), https://the-digital-reader.com/2018/02/22/internet-archive-ignores-dmca-notices/ [https://perma.cc/5844-LGJG].

87. Debra Weinstein, Note, *Defining Expeditious: Uncharted Territory of the DMCA Safe Harbor Provision*, 26 CARDOZO ARTS & ENT. L. REV. 589, 592 (2008).

88. Kimberly Buffington & Carolyn S. Toto, *The Complicated Relationship between DMCA Takedown Notices and the Word 'Expeditious,'* PILLSBURY INTERNET & SOCIAL MEDIA LAW BLOG (Jan. 19, 2016), https://www.internetandtechnologylaw.com/the-complicated-relationship-between-dmca-takedown-notices-and-the-word-expeditious/ [https://perma.cc/45BQ-W9QZ].

overwhelming number of takedown notices filed.[89] But delayed responses to takedown notices could make the removal process meaningless. This is especially true with respect to unauthorized live streaming, as demonstrated next.[90]

Moreover, from a technological aspect, it has been argued that the practice of taking down allegedly infringing material has lost its efficiency due to "the rapid repopulation of links and files on file-sharing sites, including rapid community reposting and — in some cases — automated systems for rotating links on linking sites."[91] Following the shift from individual storage to distributed provisioning, removing individual allegedly infringing files often fails to target the dynamics of content storage and access.[92] Particularly, cloud architecture is characterized by the maintenance of one or a few copies of widely used files, and then apportioning access to as many users as needed.[93] Therefore, removing a single link to an alleged infringing file may be worthless. On the other hand, if a targeted file belongs to multiple users, deleting it completely might fail to distinguish between infringing and non-infringing uses.[94] Indeed, some platforms try to address these concerns. For instance, by adopting hash-based matching to remove or prevent all publicly shared links to hash-matched files upon receiving a notice for a particular file.[95] However, because such practices may result in over-blocking of non-infringing content, not all platforms are inclined to use them. Moreover, because such practices are not required under the DMCA, they effectively remain voluntary.

Another problem with notice-and-takedown is its geographical limitations. "The actual content streamed on . . . sites may be located on a Content Delivery Network that may be owned by another ISP, in a different country to that which hosts the website[.]"[96] Consequently, "it is often very hard to ascertain who exactly is running the site, and from which country."[97] Notice-and-takedown is hence hardly capable of addressing extra-territorial unauthorized file sharing and streaming.[98] As shown henceforth, sites that employ "hardcore

89. *Government Requests to Remove Content*, GOOGLE TRANSPARENCY REPORT, *supra* note 18; https://transparencyreport.google.com/copyright/overview [https://perma.cc/H5KU-NDWJ].

90. *See infra* Section II.B.2.

91. URBAN, KARAGANIS & SCHOFIELD, *supra* note 74, at 56.

92. *Id.*

93. *Id.*

94. *Id.*

95. *Id*. at 56–57.

96. NETRESULT INTELL. PROP. PROT., UPDATE ON DIGITAL PIRACY OF SPORTING EVENTS 28 (2011), https://www.wipo.int/export/sites/www/ip-sport/en/pdf/piracy_report_2011.pdf [https://perma.cc/QHZ3-UYKL].

97. *Id.*

98. *See id.*

institutional models built on piracy" appear to be the hardest to deal with.[99] As these sites contain millions of copyrighted works, sending millions of takedown notices is probably not a viable long-term solution to stop their unauthorized dissemination.

2. Example I: Live Streaming

One major area in which the system of notice-and-takedown notably malfunctions concerns live streaming of copyrighted content. Pirated TV services around the globe, and increasingly also social media platforms, have become the prevalent channel on which to watch unauthorized streaming of live events, including sports.[100] According to Muso, a piracy data company, people made 362.7 million visits to sports piracy websites in January 2019 alone.[101] Indeed, "[g]one are the days of choppy video or untimely lagging. The pirated sports streams of today are perfect with minimal lag time, often featuring a video feed so clean and clear that it can be difficult to distinguish them from the legitimate source."[102] In August 2017, for example, VFT Solutions, which specializes in monitoring live streams in social media, reported more than 7,000 partial or full live streams of the fight between the boxers Floyd Mayweather, Jr., and Conor McGregor on social media platforms, with roughly 100 million viewers, or an average of 14,000 viewers per stream.[103] While technology experts contend that they are capable of automatically detecting hundreds or even thousands of pirated live streams during a relatively short live event, the biggest challenge is shutting down this volume of live streams in a timely manner.[104]

---

99 . URBAN, KARAGANIS & SCHOFIELD, *supra* note 74, at 62; *see also infra* Section II.B.3.

100. Nelson Granados, *World Cup Live-Streaming Piracy Thrived on Social Media Platforms*, FORBES (July 18, 2018, 8:09 AM), https://www.forbes.com/sites/nelsongranados/2018/07/18/world-cup-live-streaming-piracy-thrived-on-social-media-platforms/#504148da259a [https://perma.cc/NS6Y-WRCQ].

101. Henry Bushnell, *Inside the Complex World of Illegal Sports Streaming*, MUSO, https://www.muso.com/magazine/inside-the-complex-world-of-illegal-sports-streaming [https://perma.cc/SX6R-6JH6].

102. *OK Google — It's Time to Remove Sports Piracy Streams from Your Search Results*, CREATIVE FUTURE (July 3, 2019), https://creativefuture.org/google-sports-piracy/ [https://perma.cc/J85J-5KKB].

103. Nelson Granados, *Tens of Millions Watched Mayweather Beat McGregor on Pirate Streams*, FORBES (Aug. 28, 2017, 11:57 AM), https://www.forbes.com/sites/nelsongranados/2017/08/28/tens-of-millions-watched-mayweather-beat-mcgregor-on-illegal-streams/#5481731d79a3 [https://perma.cc/2UKE-MGEP]. Similarly, "Irdeto, a firm that provides 24x7 monitoring of global internet piracy, reported that of a regional sample of 239 illegal live streams of the fight with almost 3 million viewers, 69% were on social media channels like YouTube Live, Periscope, Facebook Live, and Twitch." *Id.*

104. *See id*.

A major problem is that the streaming world is a convoluted ecosystem whose thousands of players and internet nodes transcend thorough comprehension. It's a software developer in China, a server farm in Spain and a black-market businessman in Oklahoma. It's Reddit and YouTube, hundreds of top-level domains like '.sx' you've never heard of, and hundreds of websites you'll never see.[105]

Sending DMCA notice-and-takedown requests in these cases is hardly effective both because the streaming servers may be located outside the U.S. and because following each takedown, new streams pop up swiftly in a tiring game of whack-a-mole.[106]

Interestingly, European countries are enjoying meaningful success in addressing the problem. The English Premier League, for instance, had obtained a renewed High Court Order in 2019 that approves a blocking mechanism that will enable the blocking of a number of unidentified servers associated with infringing Premier League match footage by ISPs until the end of the 2019/20 Premier League season.[107] According to William Bush, the executive director of the Premier League, this so called "super block," which forces Internet Service Providers to disrupt or block servers hosting illegal live streams, "is proving an increasingly efficient way of blocking illegal streams at the server level[.]"[108] Yet, whether U.S. courts may harness the enforcement power of non-liable ISPs in such a way and order them to actively block their users' access to pirate streaming sites is questionable, given the longstanding restriction on injunctions against third parties.[109]

3. Example II: Copyright Infringement by Foreign Websites

Notice-and-takedown is also failing to address copyright infringement committed on foreign services. Technically, it is possible to send DMCA takedown notices to OSPs located outside the U.S. as there are many jurisdictions in which a system of notice-and-takedown exists, including Australia, China, the European Union, France, Germany, New Zealand, Singapore, South Korea, and the United Kingdom.[110] However, there are procedural and definitional

105. Bushnell, *supra* note 101.

106. *See id*.

107. *See* Andy Maxwell, *Premier League & UEFA Obtain Court Orders to Block Piracy in 2019/20*, TORRENTFREAK (July 29, 2019), https://torrentfreak.com/premier-league-uefa-obtain-court-orders-to-block-piracy-in-2019-20-190729/ [https://perma.cc/EUT7-M65S].

108. Harry Pettit, *RED CARD Premier League 'Super Block' Will Shut Down Your Illegal Online Match Streams for Entire 2019/20 Season*, THE SUN (July 30, 2019, 4:32 PM), https://www.thesun.co.uk/tech/9614609/premier-league-super-block-illegal-stream-season/ [https://perma.cc/43V6-2CXG].

109. *See infra* Section IV.A.1.

110. *See* Daniel Seng, *Comparative Analysis of the National Approaches to the Liability of Internet Intermediaries* (WORLD INTELL. PROP. ORG., Working Paper, 2010),

deviations between each system of notice-and-takedown, and therefore there is no guarantee that foreign jurisdictions, which are not bound by the DMCA, will actually comply with a DMCA takedown notice.[111]

For instance, Sci-Hub, a website that provides free access to millions of proprietary academic papers, is under a controversial attack by prominent academic publishers, such as Elsevier and the American Chemical Society ("ACS"), that claim Sci-Hub infringes their copyrights.[112] Both Elsevier and the ACS succeeded in convincing U.S. courts that Sci-Hub is effectively "The Pirate Bay of science."[113] Both complainants were awarded millions in damages and secured injunctions against the site's operator, Alexandra Elbakyan.[114] Nevertheless, Elbakyan, who is now based in Russia, has ignored rulings by U.S. courts on jurisdictional grounds, claiming Sci-Hub is not a U.S.-based company, and she is not a U.S. citizen or resident.[115]

Interestingly, the District Court for the Eastern District of Virginia, which heard the suit brought by ACS against Sci-Hub, issued an exceptionally broad injunction directed not only at Sci-Hub, but also at distinct intermediaries such as ISPs, search engines, and domain name registries that are actively associated with the Sci-Hub site, ordering them to cease facilitating access to any and all domain names and websites through which Sci-Hub engages in unlawful copyright and trademark infringement.[116] What this actually means is that the court harnessed the enforcement power of non-liable

---

https://www.wipo.int/export/sites/www/copyright/en/doc/liability_of_internet_intermediaries.pdf [https://perma.cc/UFY9-C4TG]; *see also* Jeremy de Beer & Christopher D. Clemmer, *Global Trends in Online Copyright Enforcement: A Non-Neutral Role for Network Intermediaries?*, 49 JURIMETRICS J. 375, 386 (2009). It has been argued that EU law provides a similar system of intermediary liability and hence European platforms are likely to process takedown requests by U.S. rightsholders. *See* Emerald Smith, *Lord of the Files: International Secondary Liability for Internet Service Providers*, 68 WASH. & LEE L. REV. 1555, 1574 (2011).

111. URBAN, KARAGANIS & SCHOFIELD, *supra* note 74, at 21–23.

112. *See* Ernesto Van der Sar, *Anti-Piracy Efforts Are Unlikely to Beat Sci-Hub*, TORRENTFREAK (Aug. 18, 2019), https://torrentfreak.com/anti-piracy-efforts-are-unlikely-to-beat-sci-hub/ [https://perma.cc/3Y53-7NXJ]. While Sci-Hub arguably provides unauthorized access to copyrighted material, it is also celebrated for helping the progress of science as it allows researchers to bypass expensive paywalls, access articles written by colleagues and build on them for future research.

113. Karl Bode, *'The Pirate Bay of Science' Continues to Get Attacked Around the World*, VICE (Dec. 3, 2018), https://www.vice.com/en_us/article/gy7d7j/sci-hub-and-lib-gen-continue-to-get-attacked-around-the-world [https://perma.cc/P7WP-37DQ]; *see also* Elsevier Inc. et al v. Sci-Hub et al, No. 1:15-CV-04282, at *2 (S.D.N.Y. 2015); Am. Chem. Soc'y v. John Does 1-99, No. 1:17-CV-00726, at *3 (E.D. Va. 2017).

114. *See* Elsevier Inc. v. Sci-Hub, No. 1:15-CV-04282, at *2 (S.D.N.Y. 2015); *Am. Chem. Soc'y*, No. 1:17-CV-00726, at *3.

115. *See* Quirin Schiermeier, *Pirate Research-Paper Sites Play Hide-and-Seek with Publishers*, NATURE (Dec. 4, 2015), https://www.nature.com/news/pirate-research-paper-sites-play-hide-and-seek-with-publishers-1.18876 [https://perma.cc/H2NY-SRQG].

116. *See Am. Chem. Soc'y*, No. 1:17-CV-00726.

intermediaries in an attempt to stop the infringement of ACS's lawful rights, notwithstanding the strict procedural limitations of Rule 65 of the Federal Rules of Civil Procedure.[117] While this specific injunction seems overbroad both in its scope and reach, and in the discretion it accords online intermediaries who implement it, it demonstrates the inevitable need to rely on distinct intermediaries in cases of uncooperative foreign infringers of U.S. copyrights.[118] As to the practical efficacy of this injunction, anecdotal evidence shows that following the blocking of Sci-Hub, users did report greater difficulty in accessing its services.[119] As of this writing, the site nonetheless seems to be partly available online.[120]

These examples of live streaming and copyright infringement by foreign entities demonstrate that the liability-based strategy of notice-and-takedown is ill-suited to handle online copyright infringement. Often, the involvement of distinct intermediaries in stopping the infringing activity becomes extremely crucial. Unless OSPs disable access to unauthorized live streams, they will keep popping up. And unless OSPs disable users' access to foreign infringing websites, the rights of U.S. copyright owners will probably be left with insufficient legal protection.

Holding platforms liable — either based on the theory of direct liability or on the basis of secondary liability — is an important mechanism for having harmful content removed. Nonetheless, as the representative examples discussed demonstrate, this mechanism cannot assure removal of such content. Liability theories can and do incentivize platforms to apply content moderation techniques and act voluntarily against illicit content, as explained next. Nevertheless, it is only when a court finds a platform liable for the harm caused by online content that the platform can actually be forced to remove it. Outside this limited applicability of platform liability, sufficient measures to mandate platforms active involvement in addressing illicit content are currently lacking.

---

117. According to FED. R. CIV. P. 65(b)(2)(C), the persons bound by an order include those "who are in active concert or participation with anyone described in Rule 65(d)(2)(A) or (B)." *See also* discussion *infra* at Section III.A.1.

118. *See* Maayan Perel, *Digital Remedies*, 53 BERKELEY TECH. L.J. 1 (2020); *see also* Brief of CCIA as Amicus Curiae in Support of Objections to Magistrate Judge's Proposed Findings of Fact and Recommendations at 2, 4, *Am. Chem. Soc'y*, No. 1:17-CV-00726 [hereinafter Brief of CCIA] (arguing that the broad language of the injunction could "sweep in various Neutral Service Providers, despite their having violated no laws and having no connection to this case" without giving them an opportunity to be heard as required under due process); Diana Kwon, *American Chemical Society Wins Lawsuit Against Sci-Hub*, THE SCIENTIST (Nov. 7, 2017), https://www.the-scientist.com/news-opinion/american-chemical-society-wins-lawsuit-against-sci-hub-30648 [https://perma.cc/93V8-BZ9B].

119. *See* REDDIT, *Unable to Access Sci-Hub-Need a Permanent Solution*, https://www.reddit.com/r/scihub/comments/a1fmtp/unable_to_access_scihubneed_a_permanent_solution/ [https://perma.cc/DNN3-JCQR].

120. SCI-HUB, https://sci-hubtw.tw/ [https://perma.cc/22SC-WZ2J].

III. ENFORCEMENT-BASED SPEECH REGULATION BY PLATFORMS

Platforms' involvement in the removal or blocking of illicit content can also take the form of enforcement-based actions that attempt to make platforms accountable — albeit not liable — for harmful content on their services. In other words, such removal actions are not founded on questions of liability, such as whether the platforms directly published the harmful content or otherwise contributed to its proliferation. Instead, they rely on platforms' technological ability to act promptly and remove such content, if and when requested to do so. As explained henceforth, enforcement-based removal actions could be *mandatory* or *voluntary*. By mandatory removals, I refer to content blockings or removals that either follow a specific court order directing the platform, as a third party, to remove a designated piece of content, or are based on the platforms' legal obligations under the law. By voluntary removals, I refer to enforcement-based removals that are done either as part of the independent practice of content moderation by platforms or in compliance with governmental requests to remove specific content. As explained below, while mandatory removals increasingly take place outside the U.S., within the U.S. it is mainly voluntary removals that constitute enforcement-based regulation by platforms.

A. Mandatory Removals

Mandatory removals refer to blockings or removals of *illegal* content that either follow a specific court order directing the platform, as a third party, to remove a designated piece of content, or are based on the platforms' legal obligations. Such removals, however, hardly take place within the U.S. due to its free speech jurisprudence. Despite concerns that "the real threat to free speech today comes from private entities[,]"[121] interfering with the editorial discretion of private platforms to manage the content they disseminate is seen as a violation of platforms' own First Amendment rights.[122] In other words, governmental regulation of the way platforms present users' content, either by requiring them to take down or restrict access to different types of content, constitutes state action that implicates the First Amendment. Moreover, under the state action doctrine, constitutional free speech protections generally apply only when a person is harmed by an action of the government, not a private

121. U.S. Telecomm. Ass'n v. FCC, 855 F.3d 381, 434 (D.C. Cir. 2017) (Kavanaugh, J., dissenting).

122. *See* KELLER, *supra* note 12, at 23.

party.[123] Hence, in the U.S., platforms are treated as private actors and generally cannot be forced to take down content.

Outside the U.S., however, governments increasingly oblige platforms to act expeditiously and remove illegal content. One example is the Act to Improve the Enforcement of Rights on Social Networks (NetzDG), which was adopted in Germany in 2017. The law requires platforms to delete content that is "clearly illegal" within 24 hours of a complaint being filed.[124] Equivalent initiatives were introduced in the United Kingdom and the Russian Federation.[125] Similarly, a recent proposal by the European Commission would require hosting service providers to remove terrorist content online or disable access to it within one hour of receipt of a removal order.[126]

Below, I provide several examples of mandatory content removals by platforms in the U.S. Although these examples are very particular and fairly exceptional, they indicate that mandatory enforcement-based speech regulation by platforms is not a complete stranger to the U.S. legal system. While such regulation efforts are restricted by some serious barriers, which I discuss in Parts IV and V, overcoming them is and should be possible.

1. Court Orders Directed at Platforms as Third Parties

Platforms could be obligated to block or remove illegal content if a court orders them to do so. While such obligations may be theoretically bound by the First Amendment and override governing procedural law, as I discuss in Part IV, in practice, they are nonetheless exploited not unfrequently. For instance, in several recent trademark cases, judges have ordered search engines to block access to counterfeiting sites by excluding them from their search results.[127] In fact, these orders were directed at not only search engines, but also at social media websites, including Facebook, Google+, and Twitter.[128] Importantly, in these cases, the enjoined platforms were

123. *See, e.g.*, Lloyd Corp. v. Tanner, 407 U.S. 551, 567 (1972).

124. Netzdurchsetzunggesetz [NetzDG] [Network Enforcement Act], Sept. 1, 2017 *translation at* https://germanlawarchive.iuscomp.org/?p=1245 [https://perma.cc/E4WL-Y2HE] (Ger.) [hereinafter NetzDG].

125. *See* Katharina Kaesling, *Privatizing Law Enforcement in Social Networks: A Comparative Model Analysis*, 3 ERASMUS L. REV. 151, 152–56 (2018).

126. *See* European Commission Press Release IP/18/5561, State of the Union 2018: Commission Proposes New Rules to Get Terrorist Content off the Web (Sept. 12, 2018), https://ec.europa.eu/commission/presscorner/detail/en/IP_18_5561 [https://perma.cc/H4XP-KQZG].

127. *See* Courtney Brown, *Caught in a Bind: Reassuring Judicial Authority to Bind Non-Party Search Engines under Rule 65 in Counterfeit Goods Cases*, 32 CARDOZO ARTS & ENT. L.J. 257, 259 & n.13 (2014).

128. *See id*.

not included as parties to the lawsuit, and thus their liability was not discussed or determined.[129]

Obligating online intermediaries to block access to sites selling counterfeits seems to be the best and perhaps only cure against counterfeiters operating online.[130] Stopping the online sale of counterfeit goods is extremely difficult, especially given that counterfeiters can easily register new domain names if their previous ones are seized.[131] Moreover, since counterfeiters provide false contact information to domain name registries, it is almost impossible to seize counterfeit goods.[132] Furthermore, counterfeiting assets are often located outside the jurisdiction of U.S. courts, making it even more difficult to successfully fight counterfeiters.[133] Hence, the technological control of online intermediaries over U.S. users' access to websites — including counterfeiting websites — makes them key actors in enforcing the rights of trademark owners. Without their assistance, "there may always be a way for counterfeiters to survive[.]"[134]

Another example of a court order directed at platforms as third parties relates to copyright law. In the *Sci-Hub* case discussed earlier, the District Court for the Eastern District of Virginia issued a broad injunction, not only ordering Sci-Hub (the allegedly infringing website) to stop distributing copyrighted content and infringing trademarks, but also directing all those "in active concert or participation" with Sci-Hub, "including Internet search engines, web hosting and Internet service providers, domain name registrars, and domain name registries" to cease facilitating access to any or all domain names and websites through which Sci-Hub operates.[135] Since the Sci-Hub website was operated from out of the country, and its operator, Alexandra Elbakyan, showed no intention to comply with the court's order, the assistance of U.S. intermediaries was crucial to assure the order would be enforced.[136] Unless these intermediaries prevent U.S. users from accessing the changing domains through which the Sci-Hub website operates, the rightsholders will be left with no meaningful recourse against the ongoing infringement of their intellectual property rights.

These examples suggest that orders against third party platforms seeking their cooperation in enforcing the legitimate rights of

---

129. *See id*. at 259.
130. *See id.* at 260.
131. *See id.* at 258.
132. *See id.* at 258–59.
133. *See id.* at 259.
134. *Id.* at 263.
135. Am. Chem. Soc'y v. John Does 1-99, No. 1:17-CV-00726, at *3 (E.D. Va. 2017); *see also* Brown, *supra* note 127, at 269.
136. *See* Brown, *supra* note 127, at 263; *see also* Schiermeier, *supra* note 115.

individuals, irrespective of questions of platforms' liability, are nominally granted by U.S courts. This does not necessarily mean, however, that courts have the ability to enforce them.[137] Moreover, besides the procedural difficulties that will be discussed later in Part IV, it is also important to mention the serious implications such broad orders might have on fundamental rights.[138] Blocking users' access to legitimate online content that happens to reside on the blocked website could curtail their First Amendment right to freely consume information in the marketplace of ideas.[139] Nevertheless, as I explained in a prior article, courts can apply different measures to ensure both that the grant of such orders is appropriate and proportional, and that their implementation is subject to ongoing judicial review.[140] Indeed, unlike voluntary enforcement measures, which depend on the platforms' discretion and are not subject to judicial oversight or constraints, mandatory measures could be overseen by the court.[141]

2. Platforms' Enforcement Obligations Under U.S. Law

Mandatory content removals by platforms could also be based on platforms' enforcement obligations under the law. While foreign governments, as noted earlier, are increasingly pushing toward the adoption of such legislation, the U.S. legal system is extremely restricted in regulating platforms. The Free Speech Clause of the First Amendment provides that "Congress shall make no law . . . abridging the freedom of speech," and applies to the states through the Fourteenth Amendment.[142] Generally, the government may not regulate speech "because of its message, its ideas, its subject matter, or its content."[143] For content-based regulation to be deemed

137. *See* Brown, *supra* note 127, at 268; *see also* Hassell v. Bird, 420 P.3d 776 (Cal. 2018) (vacating a lower court order, based upon a default judgment in a defamation action, which had directed Yelp, Inc., a non-party to the original suit, to take down certain consumer reviews posted on its site).

138. *See* Perel, *Digital Remedies*, *supra* note 118.

139. *Id.* at 39; Jerome A. Barron, *Access to the Press — a New First Amendment Right*, 80 HARV. L. REV. 1641, 1666–78 (1967) (discussing rights of access to the press); Jamie Kennedy, *The Right to Receive Information: The Current State of the Doctrine and the Best Application for the Future*, 35 SETON HALL L. REV. 789, 789–90 (2005); Susan Nevelow Mart, *The Right to Receive Information*, 95 L. LIBR. J. 175, 175 (2003).

140. *See* Perel, *Digital Remedies*, *supra* note 118, at 43–51.

141. *See infra* Section VI.A.

142. U.S. CONST. amend. I; *see* Cantwell v. Connecticut, 310 U.S. 296, 303 (1940) ("The fundamental concept of liberty embodied in th[e Fourteenth] Amendment embraces the liberties guaranteed by the First Amendment . . . The Fourteenth Amendment has rendered the legislatures of the states as incompetent as Congress to enact [laws in contradiction of the First Amendment].").

143. Police Dep't of Chi. v. Mosley, 408 U.S. 92, 95 (1972). *See, e.g.*, Erznoznik v. City of Jacksonville, 422 U.S. 205 (1975); First Nat'l Bank of Bos. v. Bellotti, 435 U.S. 765 (1978); Carey v. Brown, 447 U.S. 455 (1980); Metromedia, Inc. v. City of San Diego, 453

constitutional, the government must show that the regulation "is necessary to serve a compelling state interest and is narrowly drawn to achieve that end."[144]

The Supreme Court has held that the First Amendment permits restrictions upon the content of speech in a "few limited areas," including obscenity, defamation, fraud, incitement, fighting words, and speech integral to criminal conduct.[145] The reasoning behind this interpretation was that

> "certain well-defined and narrowly limited classes of speech . . . are no essential part of any exposition of ideas, and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality."[146]

Recent decisions of the Court, however, seem to interpret these categories rather narrowly.[147]

In fact, even within these "limited areas," due to the barriers of § 230, there are "few, if any, federal or state laws that expressly govern" platforms' content-related decisions.[148] Notwithstanding public policy concerns about harmful content online, prior legislative attempts to mandate speech regulation by platforms for the benefit of the public were unsuccessful. For instance, the provisions of the CDA that prohibited the transmission of indecent or patently offensive messages to minors were struck down as unconstitutional under the First Amendment.[149] Similarly, the Court held that a federal statute

---

U.S. 490 (1981) (plurality opinion); Widmar v. Vincent, 454 U.S. 263 (1981); Regan v. Time, Inc., 468 U.S. 641 (1984).

144. Ark. Writers' Project, Inc. v. Ragland, 481 U.S. 221, 231 (1987).

145. United States v. Stevens, 559 U.S. 460, 468 (2010).

146. Chaplinsky v. New Hampshire, 315 U.S. 568, 571–72 (1942).

147. *See* United States v. Alvarez, 567 U.S. 709, 718 (2012) (plurality opinion) ("Absent from those few categories where the law allows content-based regulation of speech is any general exception to the First Amendment for false statements."); Brown v. Entm't Merchs. Ass'n, 564 U.S. 786, 792 (2011) (citing Winters v. New York, 333 U.S. 507, 517–19 (1948) (holding that the obscenity exception to the First Amendment does not cover violent speech)); *Stevens*, 559 U.S. at 472 (declining to "carve out" an exception to First Amendment protections for depictions of illegal acts of animal cruelty); Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 55 (1988) (refusing to restrict speech based on its level of "outrageousness").

148. VALERIE C. BRANNON, CONG. RSCH. SERV., R45650, FREE SPEECH AND THE REGULATION OF SOCIAL MEDIA CONTENT 16 (2019), https://fas.org/sgp/crs/misc/R45650.pdf [https://perma.cc/NK7B-NDVC].

149. Reno v. Am. C. L. Union, 521 U.S. 844, 844 (1997) (holding that such a law was overbroad because content platforms and other online intermediaries could not always determine who their audience was, and thus the law would essentially require a lowest-common denominator approach to Internet publication. The holding reserved the government's right to investigate and prosecute child pornography.).

prohibiting "sexually explicit images that appear to depict minors" was found unconstitutionally overbroad.[150] Other legislative attempts, especially the proposed Stop Online Piracy Act, which sought to give the U.S. Attorney General the ability to obtain injunctions against "foreign infringing sites," and the proposed Protect IP Act, which sought to enable the government and rightsholders to combat infringing websites, were abandoned before enactment. [151] Consequently, the regulation of online speech by platforms remains largely voluntary, governed primarily by platforms' private content moderation policies.[152]

B. Voluntary Removals

Most major platforms voluntarily take action against harmful speech, either based on their independent judgment or as a response to requests submitted by law enforcement agents. Platforms remove illicit content on their sites — albeit not being legally required to do so — for several reasons.[153] First, troubling content, such as hate speech, incitement, and misinformation, may drive away users and advertisers. Hence, to protect their economic interests in their reputation, platforms seek to maintain the appearance of respectability.[154] Second, platforms remove harmful content out of a sense of public obligation.[155] Often, they remove objectional content

---

150. Ashcroft v. Free Speech Coal., 535 U.S. 234, 239–40 (2002) (holding that, notwithstanding the government's power to prosecute actual cases of child pornography, the statute violated the First Amendment because it "proscribe[d] a significant universe of speech that is neither obscene . . . nor child pornography").

151. H.R. 3261, 112th Cong. § 102(b)(5) (1st Sess. 2011); S. 968, 112th Cong. (1st Sess. 2011); *see also* Mike Masnick, *An Updated Analysis: Why SOPA & PIPA Are A Bad Idea, Dangerous & Unnecessary*, TECHDIRT (Jan. 18, 2012, 7:32 AM), http://www.techdirt.com/articles/20120117/23002717445/updated-analysis-why-sopa-pipa-arebad-idea-dangerous-unnecessary.shtml [https://perma.cc/E4JD-9XLW]

152. *See* Klonick, *supra* note 23, at 1630–58 (2018) (describing some of these policies).

153. *See* Jason Koebler & Joseph Cox, *The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People*, VICE (Aug. 23, 2018, 5:57 PM), https://www.vice.com/en_au/article/xwk9zd/how-facebook-content-moderation-works [https://perma.cc/9N5E-CP5J] (statement of Sarah T. Roberts) ("The fundamental reason for content moderation — its root reason for existing — goes quite simply to the issue of brand protection and liability mitigation . . . It is ultimately and fundamentally in the service of the platforms themselves. It's the gatekeeping mechanisms the platforms use to control the nature of the user-generated content that flows over their branded spaces.").

154. *See, e.g.*, Gillespie, *supra* note 14; Niva Elkin-Koren & Maayan Perel, *Separation of Functions for AI: Restraining Speech Regulation by Online Platforms*, 24 LEWIS & CLARK L. REV. 857 (2020).

155. *See, e.g.*, Andrew Hutchinson, *Facebook Announces New Policy to Crackdown on Manipulated Media*, SOCIALMEDIATODAY (Jan. 7, 2020), https://www.socialmediatoday.com/news/facebook-announces-new-policy-to-crackdown-on-manipulated-media/569907/ [https://perma.cc/V2A2-R3H5].

as a response to public outcry.[156] Third and finally, platforms remove potentially illicit content to mitigate liability risks.[157] Nevertheless, as described in Section III.A, governmental requests to remove presumably illicit content do not guarantee its removal. Additional procedures for mandatory removals, such as those proposed in Parts IV and V, should be adopted to better protect our online public sphere.

C. Content Moderation

Section 230 also affords protection for "Good Samaritan" content moderation.[158] The idea behind this protection was to assure that the safe harbor does not discourage platforms from voluntarily taking action against harmful content disseminated through their platforms, fearing a heightened threshold of liability.[159] And so, it appears that platforms have expanded their voluntary content moderation practices as a result of § 230.

Content moderation by humans is extremely common,[160] but so is moderation by Artificial Intelligence ("AI"). Indeed, with the growth in the amount of content posted online, as well as the public and regulatory pressure on platforms to protect users and expeditiously remove illicit content, it has become almost impossible for platforms

---

156. *See* Ina Fried, *Youtube Tightens Hate Speech Policies*, AXIOS (June 5, 2019), https://www.axios.com/youtube-tightens-hate-speech-policies-c840d9e6-fbc7-49bc-aaf5-76f818b76190.html [https://perma.cc/37MS-AWNC]; Steve Kovach, *YouTube Says It is Banning Supremacist Videos*, CNBC (June 5, 2019, 12:11 PM), https://www.cnbc.com/2019/06/05/youtube-to-ban-supremacist-content.html [https://perma.cc/38GA-2DK8]; Chris Stokel-Walker, *YouTube's Plan to Fix Hate Speech Failed Before It Even Started*, WIRED (June 6, 2019), https://www.wired.co.uk/article/youtube-steven-crowder-ban-hate-speech [https://perma.cc/E5QE-JKQ2].

157. I use the term "potentially" because this Part talks about out-of-court removals of objectionable content, so there is no legal determination about the legality of the content. From a global perspective, recent regulatory efforts have expanded the potential liability of online platforms for potentially harmful content on their websites, further pushing platforms to engage in content moderation. For instance, the German government has introduced the Network Enforcement Act ("NetzDG"), which requires major social network providers to delete unlawful content within a short timeframe after a complaint has been filed. NetzDG, *supra* note 124. The EU's new Copyright in the Digital Single Market Directive assigns greater responsibility to platforms to monitor and screen user content uploads. Council Directive 2019/790, 2019 O.J. (L 130). Thus, under certain circumstances, platforms are now practically forced to adopt content-moderation strategies.

158. 47 U.S.C. § 230(c).

159. 47 U.S.C. § 230(c)(2) ("No [platform] . . . shall be held liable on account of — (A) any action voluntarily taken in good faith to restrict access to or availability of material that the . . . [platform] considers . . . [in any way] objectionable, whether or not such material is constitutionally protected.").

160. Brittan Heller, *What Mark Zuckerberg Gets Wrong — and Right — About Hate Speech*, WIRED (May 2, 2018, 8:00 AM), https://www.wired.com/story/what-mark-zuckerberg-gets-wrongand-rightabout-hate-speech/ [https://perma.cc/6AP8-SLXH]; Sebastian Felix Schwemer, *Trusted Notifiers and the Privatization of Online Enforcement*, 35 COMPUT. L. & SEC. REV. 6, 6 (2019).

to rely exclusively on human reviewers for content moderation purposes. For instance, in relation to terrorist content, Facebook has recently admitted that 99% of the terrorist content it removes is flagged by AI-based systems before anyone on Facebook's services reports it.[161] YouTube has announced that it is using AI to spot extremist content, and that more than 83% of the videos it deleted were flagged by AI and three quarters of those were deleted before they got any views.[162]

Content moderation by platforms can make our public sphere a safer place. At the same time, however, it can be over-protective, silencing legitimate or marginalized speech.[163] Indeed, this is largely due to the scale of online content. If moderation was once driven by devoted community management that aimed to protect the special values of members of various communities on a case-by-case basis,[164] today content moderation is broad and generalized. Similar cases are decided similarly, possibly ignoring the specific context of the speech.[165] False positives are frequent, but so are false negatives.[166] Indeed, plenty of undesired content remains online "for days, or years, because of the sheer challenge of policing

---

161. *Facebook's AI Wipes Terrorism-Related Posts*, BBC (Nov. 29, 2017), https://www.bbc.com/news/technology-42158045 [https://perma.cc/DBQ2-QZLX]; *see also* Emily Dreyfuss, *Facebook Streams a Murder, and Must Now Face Itself*, WIRED (Apr. 16, 2017, 9:26 PM), https://www.wired.com/2017/04/facebook-live-murder-steve-stephens/ [https://perma.cc/U536-WJS4].

162. Kate O'Flaherty, *YouTube Keeps Deleting Evidence of Syrian Chemical Weapon Attacks*, WIRED (June 26, 2018), https://www.wired.co.uk/article/chemical-weapons-in-syria-youtube-algorithm-delete-video [https://perma.cc/7477-5YXW]; David Meyer, *AI Is Now YouTube's Biggest Weapon Against the Spread of Offensive Videos*, FORTUNE (Apr. 24, 2018, 5:56 AM), https://fortune.com/2018/04/24/youtube-machine-learning-content-removal/ [https://perma.cc/45UF-J77V]; Susan Wojcicki, *Expanding Our Work Against Abuse of Our Platform*, YOUTUBE OFF. BLOG (Dec. 5, 2017), https://blog.youtube/news-and-events/expanding-our-work-against-abuse-of-our [https://perma.cc/H24S-7ZXX]; The YouTube Team, *An Update on Our Commitment to Fight Violent Extremist Content Online*, YOUTUBE OFF. BLOG (Oct. 17, 2017), https://blog.youtube/news-and-events/an-update-on-our-commitment-to-fight [https://perma.cc/UL64-3E2K].

163. Corynne Macsherry, *Platform Censorship: Lessons From the Copyright Wars*, ELECT. FRONTIER FOUND. (Sept. 26, 2018), https://www.eff.org/deeplinks/2018/09/platform-censorship-lessons-copyright-wars [https://perma.cc/GLF4-4TXD]; Queenie Wong, *Is Facebook Censoring Conservatives or Is Moderating Just Too Hard?*, CNET (Oct. 29, 2019), https://www.cnet.com/features/is-facebook-censoring-conservatives-or-is-moderating-just-too-hard/ [https://perma.cc/6G8P-J4DX].

164. Ashcroft v. Free Speech Coal., 535 U.S. 234, 240 (2002) (holding that notwithstanding the government's power to prosecute actual case of child pornography, the statute violated the First Amendment because it "proscribe[d] a significant universe of speech that is neither obscene . . . nor child pornography").

165. Ben Depoorter & Robert Kirk Walke, *Copyright False Positives*, 89 NOTRE DAME L. REV. 319, 320–21 (2013).

166. *Id.*; Daphne Keller, *Empirical Evidence of "Over-Removal" by Internet Companies Under Intermediary Liability Laws*, CTR. FOR INTERNET & SOC'Y (Oct. 12, 2015, 8:23 AM), http://cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws [https://perma.cc/QX8F-Z4AB].

platforms."[167] This suggests that voluntary content moderation is simply insufficient to guarantee removal of illicit content.

Another reason why voluntary content moderation cannot guarantee the removal of illicit content relates to privatization.[168] It is private platforms that elaborate rules and systems to resolve conflicts between preserving free expression and regulating harmful speech.[169] Many platforms opt to make content moderation decisions based on their internal terms of service.[170] However, platforms do not only set the "laws of flagging," they also apply and enforce them by automated means.[171] As I have shown elsewhere, when these internal guidelines about what is illegitimate content are deployed by a complicated system of AI, removal decisions can be tainted by the platforms' private economic interest in maintaining controversial content.[172] Hence, there is a genuine risk that, instead of removing potentially harmful content, platforms will actually encourage its dissemination.[173]

1. Governmental Requests

Voluntary removal of harmful content by platforms can also rely on governmental removal requests. Various platforms allow governmental agents and other authorized reporters to file requests to

167. Gillespie, *supra* note 14, at 16.

168. Perel & Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, *supra* note 81, at 481; Schwemer, *supra* note 160, at 6; Wendy Seltzer, *Free Speech Unmoored in Copyright's Safe Harbor: Chilling Effects of the DMCA on the First Amendment*, 24 HARV. J.L. & TECH. 171, 179 (2010).

169. *See, e.g.*, Klonick, *supra* note 23, at 1630–58.

170. BEN WAGNER, GLOBAL FREE EXPRESSION: GOVERNING THE BOUNDARIES OF INTERNET CONTENT 128 (2016).

171. Thomas E. Kadri & Kate Klonick, *Facebook v. Sullivan: Public Figures and Newsworthiness in Online Speech*, 93 S. CAL. L. REV. 37, 59 (2019).

172. Elkin-Koren & Perel, *Separation of Functions*, *supra* note 154, at 43–48.

173. A study at the Harvard' Berkman Klein Center for Internet and Society found that YouTube's recommendation system is suggesting home videos of partially clothed children to users, sometimes after these users watched sexually explicit content. While each family home video on its own is perfectly innocent, when grouped with sexually explicit materials, its meaning might change. *See* Max Fisher & Amanda Taub, *On YouTube's Digital Playground, an Open Gate for Pedophiles*, N.Y. TIMES (Jun. 3, 2019), https://nyti.ms/2Is2PX5 [https://perma.cc/3H9M-98M6]; Jonas Kaiser & Yasodara Córdova, *On YouTube's Digital Playground*, BERKMAN KLEIN CTR. (Jun 3, 2019), https://cyber.harvard.edu/story/2019-06/youtubes-digital-playground [https://perma.cc/5DEL-2ZN5]; *see also* Kerry Jones, Kelsey Libert & Kristin Tynski, *The Emotional Combinations that Make Stories Go Viral*, HARV. BUS. REV. (May 23, 2016), https://hbr.org/2016/05/research-the-link-between-feeling-in-control-and-viral-content [https://perma.cc/56UR-CBSR] (explaining that posts go viral when they maximize feelings of arousal and dominance and that marketers can take advantage of this fact in promoting content).

remove allegedly illicit content from their services.[174] According to the Google Transparency Report, content removals in the U.S. mostly include de-listings due to copyright infringement, YouTube Community Guidelines enforcement, defamation, or violation of local laws prohibiting hate speech or adult content.[175] There has been exponential growth in the filings of governmental requests by U.S. officials. For instance, according to the Twitter Transparency Center, Twitter received two governmental removal requests between July and December 2012, 26 requests between July and December 2014, and 100 requests between July and December 2016.[176]

Nevertheless, platforms are not legally bound by such removal requests, so they can elect to partially or completely decline them.[177] As stated by Google Transparency Report, governmental requests — even those including court orders — "do not compel Google to take any action."[178] These requests may result from a dispute with a third party. They are submitted by the requesting user as evidence to support her claim that Google should remove the content. According to Google Transparency Report, 60% of the requests from governmental agencies or law enforcement agents were partially or completely acted upon.[179] This suggests that many governmental requests — 40% of total requests — are practically declined.

Government regulation of platforms' speech regulation infrastructure has been named the "new school" of online governance.[180] Platforms' engagement in addressing illicit and harmful content online is critical "to protect the values of a democratic culture and the ability of individuals to participate in the public sphere."[181] Nevertheless, the discussion so far has shown that speech regulation by platforms is still far from maximizing its full potential. Liability-based speech regulation does not fit where platforms are not directly or contributorily involved in the publication or dissemination of harmful content. Enforcement-based speech regulation is almost exclusively voluntary, providing no assurance

---

174. *See, e.g.*, *Government Requests to Remove Content*, GOOGLE TRANSPARENCY REPORT, *supra* note 18; TWITTER TRANSPARENCY, *Removal Requests*, *supra* note 18; *Government Requests to Remove Content*, REDDIT TRANSPARENCY REPORT 2019, https://www.redditinc.com/policies/transparency-report-2019 [https://perma.cc/CR9Q-LD9U].

175. GOOGLE TRANSPARENCY REPORT, *supra* note 18.

176. *United States*, TWITTER TRANSPARENCY, https://transparency.twitter.com/en/countries/us.html [https://perma.cc/8427-BZVH].

177. GOOGLE TRANSPARENCY REPORT, *supra* note 18.

178. *Id.*

179. *Removal Requests by Country/Region*, GOOGLE TRANSPARENCY REPORT, *supra* note 18.

180. *See generally* Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149 (2018).

181. *Id.* at 1152.

that harmful content will actually be removed. Accordingly, in the next two Parts, I attempt to set the legal foundation for making enforcement-based regulation by platforms mandatory, irrespective of liability. Advancing two procedural fixes to the current legal system — first, a proposed interpretive approach that would enable courts to enjoin platforms as non-parties in civil suits; and, second, a proposed statutory procedure that will enable designated law enforcement agents to file removal requests of illegal content with the court — I discuss the barriers that could interfere with these fixes and explain how they could be overcome.

IV. SPEECH REGULATION BY NON-LIABLE PLATFORMS #1: ENJOINING PLATFORMS AS NON-PARTIES IN CIVIL SUITS

Platforms govern the infrastructure for online speech.[182] They have the technological capacity to control, limit, and censor speech.[183] Nonetheless, U.S. courts currently treat non-liable platforms as distinct, private entities that are generally located outside the reach of their injunctive power. That is true, even if only these platforms are able to stop, or at least minimize dramatically, the harms caused by tortious online content.

To fix this, I recommend confirming courts' authority to issue injunctions against non-liable platforms in civil disputes through statutory interpretation. Such injunctions would not be based on any cause of action which seeks to hold the platform liable for the harms caused to the plaintiff, but would instead seek to control the perpetuation of content declared by the court to be tortious.[184] This authority should be employed in appropriate cases, for instance, when the tortfeasor cannot be found or when he is operating from outside the jurisdiction of the court.[185] If, for instance, there is a pirate website operated by an unknown person offering live streaming of the Super Bowl, the court should be able to issue an injunction against U.S. ISPs and force them to block U.S. users' access to this pirate website until the end of the event. In the U.K., for instance, the injunction issued in the English Premier League's favor, ordering ISPs to block servers associated with infringing Premier League match footage until the end of the 2019/20 Premier League season, was proven to be highly effective.[186] Nevertheless, there are several legal barriers that must be addressed in order to assure courts can effectively enjoin non-liable platforms when necessary.

---

182. *Id*. at 1153.
183. *Id*.
184. *See* Hassell v. Bird, 247 Cal. App. 4th 1336, 1361 (2016).
185. *See supra* Section II.B.
186. Maxwell, *supra* note 107; Pettit, *supra* note 108.

A. Procedural Due Process

1. The Barriers

Basic principles of due process limit the power of courts to bind third parties, who are, effectively, "strangers to the litigation."[187] Due process of law in judicial proceedings primarily yields "the opportunity to be heard" for holding "one bound by the judgment who has not had such opportunity is contrary to the first principles of justice."[188] Hence, only a third party that deliberately works "with or for" a party to subvert the injunction can be so bound.[189]

These basic principles are embedded in the Federal Rules of Civil Procedure. Specifically, Rule 65 restricts injunctions to the parties, their officers, agents, servants, employees, and attorneys, and "other persons who are in active concert or participation" with the parties.[190] Accordingly, a third party should have a "close alliance with the enjoined defendant" before it can be bound.[191] Cases that have interpreted the rule have held non-parties liable under injunctions "when the [non-party] has aided or abetted a party in the violation of the injunction."[192]

Based on these principles, Google has argued that distinct online intermediaries cannot be required to remove content created and published by others.[193] Specifically, in *Blockowicz v. Williams*,[194] the Seventh Circuit considered whether Ripoff Report, a platform hosting a review found to be defamatory, could be held in contempt for ignoring the injunction issued by the court, which ordered the review to be removed.[195] The court held that it lacked the authority to extend the injunction to Ripoff Report because it was not proven that Ripoff Report had aided or abetted the defendant's violation of the

187. Philip Morris USA, Inc. v. Williams, 549 U.S. 346, 353 (2007); Hansberry v. Lee, 311 U.S. 32, 40 (1940) ("It is a principle of general application in Anglo-American jurisprudence that one is not bound by a judgment *in personam* in a litigation in which he is not designated as a party or to which he has not been made a party by service of process.") (citing Pennoyer v. Neff, 95 U.S. 714 (1878); 1 FREEMAN ON JUDGMENTS (5th ed. 1925), § 407).

188. Baker v. Baker, Eccles & Co., 242 U.S. 394, 403 (1917).

189. People v. Conrad, 55 Cal. App. 4th 896, 903 (1997).

190. FED. R. CIV. P. 65(d)(2).

191. Microsystems Software, Inc. v. Scandinavia Online AB, 226 F.3d 35, 43 (1st Cir. 2000).

192. Herrlein v. Kanakis, 526 F.2d 252, 254 (7th Cir. 1975).

193. Amicus Curiae Brief of Google Inc. in Support of Yelp Inc. at 11, Hassell v. Bird, 47 Cal. App. 4th 1336 (2016) ("These principles ensure that [non-party] intermediaries are not transformed into deputies, required under pain of contempt to censor material created by others.").

194. 630 F.3d 563 (7th Cir. 2010).

195. *Id.* at 569.

injunction.[196] Other than hosting the defamatory material, Ripoff Report did nothing, and this, according to the court, was not enough to justify enjoining it. Notwithstanding that the platform was technologically capable of removing the material, the court did not find that its failure to do so constituted "aiding and abetting."[197] According to one interpretation of the *Blockowicz* holding, "where an online service provider simply declines to take action called for by the court order — such as leaving content up, not removing information from its search results, or otherwise continuing to provide a general service to a party — that is not the kind of close concerted action required to subject it to contempt."[198]

Similarly, in *Hassell v. Bird*,[199] the Supreme Court of California considered whether the plaintiff's remedies could be extended through an injunction beyond the defendant to a third-party platform.[200] That case considered a bad review posted by Bird on Yelp criticizing the services of the law firm Hassell Law Group.[201] The law firm sued Bird for defamation and won a default judgment; the San Francisco Superior Court ordered Bird to remove every defamatory online review about the plaintiffs.[202] Additionally, it issued a second order against Yelp to remove the defamatory posts.[203] Yelp moved to set aside the default judgment, arguing that forcing it to comply with an injunction constituted the type of liability barred by the CDA.[204] The Superior Court disagreed, reasoning that Yelp aided and abetted Bird by highlighting at least one of her posts as a "recommended review."[205] Subsequently, the California Court of Appeals agreed with the Superior Court, finding that the CDA does not prevent a court from "directing an Internet service provider to comply with a judgment which enjoins the originator of defamatory statements posted on the service provider's Web site."[206]

The Supreme Court of California, however, reversed the Court of Appeals and ruled that the CDA safe harbor prevented the courts from ordering Yelp to remove the defamatory review.[207] Interestingly, this decision "implied that 'liability' under the CDA encompasses all legal

196. *Id*. at 568.
197. *Id*.
198. Amicus Curiae Brief of Google Inc. in Support of Yelp Inc. at 28, *Hassell*, 47 Cal. App. 4th 1336 (citing *Blockowicz*, 630 F.3d, at 568–69).
199. 420 P.3d 776 (Cal. 2018)
200. *Id.* at 782.
201. *Id*. at 778–80.
202. *Id*. at 780.
203. *Id*. at 781.
204. *Id*.
205. Order Denying Yelp's Motion to Set Aside and Vacate Judgment, Hassell v. Bird, 2014 WL 12577620, at *2 (Cal. Super. Sept. 29, 2014).
206. Hassell v. Bird, 247 Cal. App. 4th 1336, 1363 (2016).
207. Hassell v. Bird, 420 P.3d 776, 793 (Cal. 2018).

obligations, including injunctions."[208] Hence, although the Supreme Court of California acknowledged that "as a general rule, when an injunction has been obtained, certain nonparties may be required to comply with its terms," it held that § 230 precludes the application of this rule to platforms who did not themselves publish the controversial content.[209] Especially relevant, the concurring opinion of Justice Kruger agreed with Hassell's appeal primarily on due process grounds. Reasoning that common law principles ordinarily only allow for an injunction against non-parties "through whom the enjoined party may act, such as agents, servants, employees, aiders, abettors, etc.,"[210] she concluded that binding Yelp without giving it the prior opportunity to defend itself violated due process.[211]

Accordingly, it seems like the due process barrier in relation to enjoining non-liable platforms encompasses three sub-hurdles: First, as a preliminary matter, using injunctions to enjoin non-liable platforms could be viewed as holding them liable, and this might violate § 230. Second, non-liable platforms are not acting in concert with the publisher of the illegal content and therefore, they cannot be enjoined. Third, binding platforms without giving them prior notice and an opportunity to be heard may violate due process. Could these hurdles be overcome to assure platforms' cooperation in combatting unlawful online activity?

2. Possible Solutions

The first hurdle set forth above could be overcome. Indeed, the contention that injunctions impose liability on platforms in contradiction to § 230 is a matter of judicial interpretation. A different interpretation would argue that to the extent that injunctions against platforms are not based on any cause of action which seeks to hold them liable, they are not bound by the CDA.[212] Indeed, Justice Liu, in a dissent filed in *Hassell v. Bird*, argued that CDA immunity did not apply to Yelp because the injunction did not impose liability on Yelp for its role as speaker or publisher of third-party content, and in fact, never determined whether Yelp's decision to post the content was

208. Sara Gold, *When Policing Social Media Becomes a 'Hassell,'* 55 CAL. W. L. REV. 445, 453 (2019).

209. *Hassell*, 420 P.3d at 789.

210. *Id.* at 795 (Kruger, J., concurring) (citation omitted).

211. *Id.* at 802.

212. Hassell v. Bird, 203 Cal. Rptr. 3d 203, 226–27 (Cal. Ct. App. 2016), *rev'd*, 420 P.3d 776 (Cal. 2018); Brief for Erwin Chemerinsky et al., as Amici Curiae in Favor of Respondents at 1–3, *Hassell*, 420 P.3d 776.

legal.[213] According to Justice Liu, § 230 was meant to ensure that website operators like Yelp do not have to incur the time or expense of litigation, but Yelp did not have these burdens in this case.[214] Additionally, § 230 purported to eliminate the pressure for website operators to decide in advance whether a statement may be "potentially defamatory," or else face legal responsibility.[215] This problem too did not exist in this case because a default judgment had already found the reviews defamatory.[216] As the second dissenter, Justice Cuéllar, concluded, while § 230 may bar injunctive relief against interactive websites who are defendants, it has nothing to do with asking Yelp to facilitate compliance with a valid court order.[217]

As to the second hurdle — whether a third-party platform is acting in concert with the publisher of the illegal content — courts should conduct a factual inquiry into the circumstances of each specific case. In *Hassell v. Bird*, for instance, it was plausible to argue that "even if Yelp was not Bird's agent or servant," it did act through Yelp because Yelp "formats the reviews, makes the reviews searchable, and aggregates reviews of each business into a rating from one to five stars."[218] Therefore, in a sense "it was Bird's defamation of Hassell, facilitated by Yelp's willing and active participation, that the trial court sought to enjoin."[219]

In other circumstances, however, it might be harder to overcome this hurdle. For instance, it would be difficult to enjoin ISPs in copyright infringement cases and force them to block users' access to pirate websites on the ground that they act in concert with the operator of the pirate website. Indeed, ISPs are merely technological "pipes," which control the infrastructure through which content is delivered to users. Unlike hosting services like Yelp, some ISPs arguably exercise no discretion in choosing which content to disseminate. Indeed, quite often they operate as "neutral service providers."[220]

Still, however, there are instances where only these distinct players can stop the infringement.[221] To authorize courts to enjoin such intermediaries, it might be necessary, in specific cases, to interpret Rule 65 to bind third party intermediaries who *enable* the infringing activity.[222] Allowing the replacement of the rough

213. *Hassell*, 420 P.3d at 803–05 (Liu, J., dissenting) (contending that § 230 should not apply to Yelp because it was not a named defendant and that requiring websites' cooperation is necessary to combat unlawful online activity).

214. *Id.* at 802–03.

215. *Id.*

216. Gold, *supra* note 208, at 456 n.65.

217. *Id.* at 456; *Hassell*, 420 P.3d at 812 (Cuéllar, J., dissenting).

218. *Hassell*, 420 P.3d at 804–05 (Liu, J., dissenting).

219. *Id.* at 805 (Liu, J., dissenting).

220. *See generally* Brief of CCIA, *supra* note 118, at 1.

221. *See* Brown, *supra* note 27, at 276.

222. *Id*. at 279.

requirement of "participation" with a lenient requirement of "enablement" would ensure illegal content online is not left unaddressed. Nevertheless, such a broad interpretation of Rule 65 should be used only in circumstances where there are no other available measures "that are adequate and less burdensome on the third parties."[223] Accordingly, when a foreign pirate website keeps infringing copyrights held by U.S. rightsholders, for instance, and the site's operator ignores the court's orders — continuing to engage in copyright infringement — a court should be authorized to obligate U.S.-based ISPs to block the infringing website. Similarly, if a foreign website provides an unauthorized live stream of the Super Bowl, those holding the exclusive right to transmit the event should be able to request a court injunction ordering U.S.-based ISPs to block U.S. users' access to that stream.

When deciding if a specific case justifies applying this broad interpretation of Rule 65, courts should balance between the need to stop the flood of illicit content online, such as defamation, revenge pornography, and intellectual property infringement, that might have serious emotional consequences for plaintiffs, and the risk of circumventing freedom of expression.[224] Hence, it is important, for instance, to consider whether the injunction was issued in a default judgment, without hearing and considering the defenses of the alleged direct infringer.[225] If it was, then it is important that the court makes extra efforts to compensate for the lack of an adversarial process and assure that legitimate content is not being suppressed. Another important consideration relates to the specific design of the service hosting the illegal content. Specifically, there are websites that cannot easily remove content once it is posted, such as blog sites generated by WordPress.[226] Moreover, courts should also consider *how* they intend the injunction to be implemented.[227] As I showed in another article, the technological implementation of injunctions is critical due to the way it affects the ultimate scope of the injunction.[228] Therefore, when appropriate, courts should prefer flexible injunctions that leave room for ex-post revision and are limited in their duration.[229]

---

223. United States v. Regan, 858 F.2d 115, 121 (2d Cir. 1988).

224. *See Hassell*, 420 P.3d at 806–07 (Cuéllar, J., dissenting); *see also* Gold, *supra* note 208, at 457.

225. *See* Perel, *Digital Remedies*, *supra* note 118, at 49.

226. Gold, *supra* note 208, at 462.

227. *See infra* Section VI.B.

228. *See generally* Perel, *Digital Remedies*, *supra* note 118.

229. *Id.* This is especially relevant to website blocking injunctions directed to ISPs because of technological nature of blocking. As I explained, blockings risk proscribing legitimate speech which might reside on the same IP address as the blocked website. Moreover, blocking orders are especially susceptible to circumvention, both from users who can disclose their real "online identity" and from infringers that can move their infringing content to alternative services and bypass the blocking. Therefore, it is extremely important

As to the third hurdle, it is possible to require plaintiffs to notify third parties of requested injunctions.[230] Such a notification could provide the relevant third party with the opportunity to prepare a defense to the injunction and present it to the court during the hearing. In this respect, it is important to emphasize that a request to enjoin a third-party platform (or intermediary) could only be submitted *after* a final judgment on the legality of the content has been rendered and the direct wrongdoer is ignoring it. Otherwise, the plaintiff will not be able to show that there are no other available measures to remedy her injury.[231]

In practice, platforms may not be able to defend against every such request to enjoin them in court, and this may raise a "selective defense" problem.[232] Indeed, if platforms' intervention in these types of cases were to become routine, they "would have to pick and choose" which cases and which requests to object to.[233] Nonetheless, as noted, requests to enjoin third party platforms should only be allowed in specific problematic cases, where the direct infringer fails to comply with the court's ruling and remove the content.

Moreover, the proposed injunctions spare much of the hassle inherent in determining content legitimacy.[234] Instead of processing content removal requests on a voluntary basis, and exercising lawmaking powers, platforms could rely on official court rulings on the legality of the content.[235] Removing content held illegal by a court is in line with the business interests of platforms striving to build brand recognition and increase users' trust. Indeed, illegal content may drive away users and advertisers.[236] Failure to remove content by

---

to assure periodically that website blockings remain relevant and effective to achieve their intended purpose.

230 . Such a requirement was recently adopted under the new Israeli Copyright legislation, which established a new legal process that allows rightsholders to file a request to the court to order third party intermediaries to block access to allegedly infringing websites. The new procedure requires the rightsholder to notify the platform and provide it with an opportunity to oppose the proposed injunction. *See* Yehuda Neubaer, *The New Israeli Online Anti-Piracy Copyright Reform Explained*, IAM (Oct. 23, 2019), https://www.iam-media.com/new-israeli-online-anti-piracy-copyright-reform-explained [https://perma.cc/J4YH-G5YG].

231. *See* United States v. Regan, 858 F.2d 115, 121 (2d Cir. 1988).

232. Gold, *supra* note 208, at 463.

233. *See id.* (discussing the selective defense problem for websites).

234. Evelyn Douek, *Finally, Facebook Put Someone in Charge*, ATLANTIC (Sept. 19, 2019), https://www.theatlantic.com/ideas/archive/2019/09/facebook-outsources-tough-decisions-speech/598249/ [https://perma.cc/S6JQ-GTB2] ("Facebook is setting up its Oversight Board because, as founder Mark Zuckerberg wrote, 'private companies should not be making so many important decisions about speech on our own.'").

235. Gold, *supra* note 208, at 463; *see also supra* Section IV.A.

236. Balkin, *Free Speech Is A Triangle*, *supra* note 21; Citron & Norton, *supra* note 21, at 1454–55 (2011).

platforms may lead to public outrage.[237] Accordingly, only where the platform believes that the court's judgment on the legality of the content was wrong should the platform have an interest in opposing the injunction. Finally, even if third party platforms eventually elect not to appear in court, due process is still satisfied, for it is about giving notice and *opportunity* to be heard.[238]

B. Prior Restraint on Speech

1. The Barrier

Enjoining third party platforms and requesting they remove online speech may also conflict with the Doctrine of Prior Restraint, which applies only to government actors.[239] This doctrine has long been understood to forbid the implementation of regulations which prevent the publication of speech prior to its distribution, including orders to remove an expression that has already been published but before it was judicially reviewed.[240] This even applies to types of expressions — such as obscenity — that are not fully protected by the First Amendment.[241] Accordingly, traditional restrictions of speech ordinarily should only be enforced by imposing ex post criminal or civil sanctions.[242] It has been argued that the main rationale for this doctrine is "the desire to prevent the chilling of speech by censorship or similar means and to ensure that all expressions are included in the marketplace of ideas."[243]

In practice, however, the Supreme Court has occasionally approved prior restraint.[244] As Bendor and Tamir explain:

---

237. *See, e.g.*, *Gatekeepers or Censors? How Tech Manages Online Speech*, N.Y. TIMES (Aug. 7, 2018), https://nyti.ms/2OkOjS6 [https://perma.cc/F4PN-LT4Y].

238. Kenneth Culp Davis, *The Requirement of Opportunity to be Heard in the Administrative Process*, 51 YALE L.J. 1093, 1093 (1942).

239. Balkin, *Free Speech is a Triangle*, *supra* note 21, at 2017–19; Klonick, *supra* note 23, at 1609.

240. Alexander v. United States, 509 U.S. 544, 550 (1993) (quoting M. NIMMER, NIMMER ON FREEDOM OF SPEECH § 4.03, at 4–14 (1984)) (defining prior restraints as "administrative and judicial orders forbidding certain communications when issued in advance of the time that such communications are to occur."); *see also* Yochai Benkler, *A Free Irresponsible Press: Wikileaks and the Battle over the Soul of the Networked Fourth Estate*, 46 HARV. C.R.-C.L. L. REV. 311, 340 (2011).

241. Conor M. Reardon, Note, *Cell Phones, Police Recording, and the Intersection of the First and Fourth Amendments*, 63 DUKE L.J. 735, 752 (2013) (discussing protection against seizures of obscene materials using Fourth Amendment doctrine to protect First Amendment values).

242. Ariel L. Bendor, *Prior Restraint, Incommensurability, and the Constitutionalism of Means*, 68 FORDHAM L. REV. 289, 291 (1999).

243. Ariel L. Bendor & Michal Tamir, *Prior Restraint in the Digital Age*, 27 WM. & MARY BILL RTS. J. 1155, 1160 (2019).

244. *See* Seattle Times Co. v. Rhinehart, 467 U.S. 20 (1984) (upholding protective order against newspaper prohibiting dissemination of information gained in pre-trial discovery);

> "Courts have approved prior restraint where the speech is deemed obscene, where a prior restraint is needed to fulfill the right to a fair trial, where the expression is part of an unprotected commercial speech, where the speech was part of a continuing course of conduct, and where the expression could endanger national security in time of emergency. Courts have also approved prior restraint in order to protect privacy, in order to prevent employment discrimination, in order to protect property, in order to regulate public forums, and in order to prevent misleading commercial expressions."[245]

Nevertheless, as I show next, it is less likely that the proposed injunctions will run afoul the Doctrine of Prior Restraint.

2. Possible Solution

To begin, prior restraint in the digital age is different from what was anticipated by the drafters of the traditional Doctrine of Prior Restraint.[246] Bendor and Tamir, for instance, explain several factors that account for this difference: the increased chilling effect of subsequent sanctions on ordinary speakers; the lesser impact of journalistic ethics on bloggers; the ease and immediacy of new media publications; the eternity of such publications; broad access to the new media; the virality of speech; and the technical ability to separate protected from unprotected speech in the digital age.[247] All these suggest that the traditional doctrine of prior restraint does not fit with the challenges raised by online media, where infringements of private entitlement, such as the right to good reputation and intellectual property rights, are far from exceptional.[248]

In practice, injunctions against online speech are actually quite prevalent. In copyright cases, for instance, preliminary injunctions targeting allegedly infringing content are granted pretty much as a matter of course, even when the defendant has engaged in creative

---

Pittsburgh Press Co. v. Pittsburgh Comm'n on Hum. Rels., 413 U.S. 376 (1973) (holding that an order prohibiting placement in sex-designated columns of advertisements for non-exempt job opportunities did not infringe the newspaper's rights); Milk Wagon Drivers Union v. Meadowmoor Dairies, Inc., 312 U.S. 287 (1941) (upholding injunction prohibiting picketing near defendant's dairy and vendor's store).

245. Bendor & Tamir, *supra* note 243, at 1161–62.

246. *Id.* at 1178.

247. *Id.* at 1180–81.

248. *Id.* at 1170.

adaptation, not just literal copying.[249] The injunctions discussed in this paper, to the contrary, are only rendered *after* the speech in question was adjudicated to be tortious, and thus do not run afoul the Doctrine of Prior Restraint.[250] As noted before, "the case law does indeed allow permanent injunctions of unprotected speech, entered after a final judicial finding that the speech is unprotected, but doesn't allow restraints entered before such a finding."[251]

Accordingly, courts should adapt the Doctrine of Prior Restraint to the challenges raised by harmful content online, and order third party platforms to remove illegal content — following full adjudication of the issue of illegality — when the direct wrongdoer fails to remove it himself.

C. Platforms' Legitimate Economic Interests

1. The Barrier

Another important issue raised by the issuance of injunctions against non-liable platforms relates to the financial costs such injunctions could potentially impose on these third parties.[252] Such injunctions effectively drag non-parties into the legal dispute, forcing them to take active enforcement actions in order to remedy the plaintiff. For instance, one possible way to implement a website blocking injunction is to apply Deep Packet Inspection ("DPI"), which uses an algorithmic filter that is located between the end-user and the Internet in general and screens all content according to specific blocking rules.[253] This blocking method is quite expensive because it depends on the development of sophisticated filtering software. [254] Inflicting high compliance costs on non-parties,

---

249. Mark A. Lemley & Eugene Volokh, *Freedom of Speech and Injunctions in Intellectual Property Cases*, 48 DUKE L.J. 147, 150 (1998).

250. *See, e.g.*, Ardia, *supra* note 22, at 51 ("[I]t is clear that a trend is emerging within both state and federal courts that permits injunctions if the speech in question was adjudged to be defamatory.").

251. Lemley & Volokh, *supra* note 249, at 175.

252. Perel, *Digital Remedies*, *supra* note 118, at 45–46; *see also* Christophe Geiger & Elena Izyumenko, *The Role of Human Rights in Copyright Enforcement Online*, 32 AM. U. INT'L L. REV. 43, 76 (2016) ("[A] fundamental right that comes to balancing in copyright website blocking cases is the freedom of access providers to conduct their business.").

253. Lukas Feiler, *Website Blocking Injunctions Under EU and U.S. Copyright Law — Slow Death of the Global Internet or Emergence of the Rule of National Copyright Law?*, 10–11 (Transatlantic Tech. L. F., Working Paper No. 13, 2012).

254. *See* OFF. OF COMMC'NS, SITE BLOCKING TO REDUCE ONLINE COPYRIGHT INFRINGEMENT: A REVIEW OF SECTIONS 17 AND 18 OF THE DIGITAL ECONOMY ACT (2011), at 40, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/78095/Ofcom_Site-Blocking_report_with_redactions_vs2.pdf [https://perma.cc/7PHT-4G7M].

especially where they are not given an opportunity to object, may raise serious due process concerns.

Nevertheless, in practice, the exploitation of legal procedures that are designed to ask third party intermediaries to take specific actions in order to facilitate the resolution of a dispute between two parties is not unusual in the U.S., notwithstanding the burden it inflicts over the complying intermediaries. For example, John Doe subpoenas allow plaintiffs to discover the identity of anonymous online speakers through third party intermediaries, like their ISP or the websites they visited.[255] Such third-party intermediaries are in no way responsible for the harm caused by an allegedly defamatory statement which was published by the anonymous speaker. Nevertheless, only the third party intermediaries may be able to provide the Internet Protocol ("IP") address associated with the anonymous publisher of the content, which is necessary to allow the plaintiff to identify the defendant and file suit.[256] Another example is the notice-and-takedown regime of the DMCA, under which online intermediaries who seek to benefit from a safe harbor have to expeditiously remove the allegedly infringing content after receiving a notification of copyright infringement submitted by the rightsholder.[257] In such cases the intermediary could be found liable if it fails to expeditiously remove the content and the rightsholder files a copyright infringement lawsuit.[258] However, at the moment of the removal, since the intermediary's liability is yet to be determined by the court, it seems fair to view the removal action taken by the intermediary as an enforcement-based action. While removal is technically voluntary, it follows a well-structured *statutory* procedure, indicating that asking intermediaries to engage in law enforcement action is not something that is new to the U.S. system.

2. Possible Solutions

Overcoming this barrier, however, should not be too difficult. First and foremost, as proposed earlier, before filing a request with the court to enjoin a third-party platform, the plaintiff will be required to notify the relevant third party, affording it with an opportunity to be heard.[259] Second, it is not at all obvious that the intermediary will have to bear the costs associated with the removal or blocking of the

255. Nathaniel Gleicher, *John Doe Subpoenas: Toward a Consistent Legal Standard*, 118 YALE L.J. 320, 325 (2008) (examining the efficacy of John Doe subpoenas and suggesting a change to the system).

256. *Id.* at 328.

257. *See supra* Section II.B.

258. 17 U.S.C. § 512 (2018).

259. *See supra* Section IV.B.2.

illegal content. Indeed, with respect to John Doe subpoenas, for instance, it is the plaintiff who is required to bear the administrative costs of obtaining the anonymous speaker's IP address from the content provider and then obtaining the anonymous speaker's identity from the ISP, as identified by the IP address.[260] Similarly, as non-U.S. courts deciding on website blocking orders often order, the costs of complying with a blocking order could be imposed, in whole or in part, on the plaintiff.[261] Nevertheless, perhaps the best solution would be to give the court issuing the injunction the discretion to decide who should bear the costs of implementing the injunction.[262] This will allow the necessary flexibility needed to address the differences in the costs of implementing different injunctions as well as applying different technological means of application.[263]

V. SPEECH REGULATION BY NON-LIABLE PLATFORMS #2: ALLOWING IN-COURT GOVERNMENTAL REMOVAL REQUESTS

As the previous Parts have demonstrated, in the U.S., speech regulation by platforms is based on liability theories. Due to First Amendment jurisprudence and the broad immunity accorded to platforms under the CDA, mandatory removals of illegal content by platforms are infrequent. In practice, too often the legal system stands powerless against harmful online content. As explained previously, such content — like counterfeits and pirated content — could be infringing upon individuals' lawful rights.[264] Harmful content can also present a threat to public safety, such as content inciting violence, child pornography, or terrorist propaganda. Accordingly, as a second fix to the current legal system, I propose passing a new statutory procedure that would allow designated law enforcement agents to file requests to remove or block illegal content with the court.[265]

Outside the realm of speech regulation, statutes that set the procedural requirement for governmental entities seeking to utilize platforms' enforcement capabilities are in fact available. One example is the Stored Communication Act ("SCA"), which established a legal

---

260 . Ronen Perry & Tal Zarsky, *Who Should be Liable for Online Anonymous Defamation?*, 82 U. CHI. L. REV. DIALOGUE 162, 166 (2015).

261. *See, e.g.*, Roadshow Films Pty Ltd. v. Telstra Corp. Ltd [2016] FCA 1503 (Austl.).

262. This solution was recently adopted by the Israeli legislature under the new revision to the Israeli Copyright Act. *See supra* note 230.

263. Perel, *Digital Remedies*, *supra* note 118, at 34–38.

264. *See supra* Section III.B.

265. Bambauer has previously presented the main attributes that should be embedded in any governmental attempt to impose "direct control" over the Net, yet he doesn't claim that such legislation should be indeed adopted. His main argument is that "hard censorship — a statute that requires the Attorney General to demonstrate that specified content is unlawful before filtering it — is preferable to soft censorship." *See* Derek E. Bambauer, *Orwell's Armchair*, 79 U. CHI. L. REV. 863, 927–36 (2012).

procedure for a governmental entity seeking action on the part of the platform.[266] The SCA controls how the government can access stored account information, such as e-mails and subscribers' billing information, through online intermediaries, such as ISPs.[267] Another example is the Obama Administration's Operation in Our Sites intellectual property enforcement effort, which allows the National Intellectual Property Rights Coordination Center, the Immigration and Customs Enforcement Office, and the Department of Justice to obtain warrants from the courts authorizing them to seize websites allegedly engaged in intellectual property violations.[268]

Within the realm of speech regulation, however, things are more complicated. A statute enabling law enforcement agents to request the removal of illegal content would face serious constitutional barriers, as explained henceforth. But these barriers are not impossible to overcome.[269] As a starting point, the new procedure would have to incorporate safeguards informed by the U.S. Constitution. Indeed, as Bambauer has previously argued, "[i]f America decides to block access to pieces of the Net . . . it should do so in a way that is open, transparent, narrowly targeted, and protective of key normative commitments such as open communication, equal treatment under the law, and due process."[270] To gain legitimacy, the new statute cannot be vague but rather openly described, transparent in what content it targets, narrow, effective and accountable.[271]

Specifically, the proposed statute would have to specify which law enforcement agents are authorized to submit a request to the court seeking a removal or blocking of illegal content. It should further afford the targeted platform with a right to be notified about the procedure in a timely manner and allow it to oppose any proposed injunction. It might also be important to allow civil rights organizations to submit their briefs, as content removals inherently affect society as a whole.[272] This is especially true when the provider of the illegal content operates outside the U.S. and thus may lack the resources or incentive to defend his rights and his viewers' rights in the U.S.

---

266. *See* 18 U.S.C. §§ 2701–2711 (2006).

267. *See What is the Stored Communication Privacy Act?*, MINC LAW, https://www.minclaw.com/legal-resource-center/what-is-the-stored-communication-privacy-act/ [https://perma.cc/X5WX-94RP].

268. For a thorough review of Operation in Our Sites, *see* Karen Kopel, *Operation Seizing Our Sites: How the Federal Government is Taking Domain Names Without Prior Notice*, 28 BERKELEY TECH. L.J. 860 (2013).

269. Balkin, *Free Speech in the Algorithmic Society*, *supra* note 180, at 1152.

270. Bambauer, *supra* note 265, at 869.

271. *Id.* at 873.

272. Klonick, *supra* note 23, at 1622–24.

Furthermore, the procedure will have to specify that it could only be used against content found to be illegal under U.S. laws, such as child pornography, threats, harassment and stalking, impersonation, extortion, solicitation, and incitement.[273] And, until there is a final adjudication on the merits of the governmental request, the content must remain online. Additionally, to assure the effectiveness of the removal or blocking, the statute must require courts to use mechanisms for ongoing oversight such as time limitation and ex post revision procedures for content blockings, to assure the injunction is not circumvented by users or content providers.[274] Finally, to facilitate accountability, the statute should guarantee an opportunity to be heard for both the content provider and the platforms, and make the courts' decisions (excluding the actual illegal content) publicly available.[275]

Accordingly, designing the technical details of the new statute in a way that will guarantee procedural due process seems rather achievable. More challenging, however, will be to situate this procedure within America's freedom of expression framework.

A. Prior Restraint

As explained previously in Part IV, it is hard to ignore the tension that exists between the doctrine of prior restraint and injunctions against illegal content.[276] Under the proposed statute, the government would be able to prevent communication "between a willing speaker and willing listeners through interdiction."[277] This sort of censorship amounts to prior restraint on speech. However, even in the U.S., where the notion of unfettered discourse is so deeply rooted, harmful content is often removed through voluntary channels.[278] Indeed, the government might enact legislation that indirectly affects speech or pressures platforms to remove objectionable content.[279] Such prior

---

273. Brandenburg v. Ohio, 395 U.S. 444, 447 (1969) (holding that the government cannot punish inflammatory speech unless that speech is "directed to inciting or producing imminent lawless action and is likely to incite or produce such action"); *see also* 18 U.S.C. § 2252A (child pornography); 18 U.S.C.§ 875 (threats); 18 U.S.C. § 2261A (harassment and stalking); 18 U.S.C. § 911 (impersonation); 18 U.S.C. §§ 871–880 (extortion); 18 U.S.C. § 373 (solicitation).

274. Perel, *Digital Remedies*, *supra* note 118, at 43, 46 (explaining that time limitations and ex post revision allow courts to limit and adjust the breath and scope of injunctions to address changes in circumstances, such as a pirate website moving to a different IP address, in a timely manner).

275. *See* Bar-Ziv & Elkin-Koren, *supra* note 80, at 495 (discussing how due process procedures facilitate accountability in automated copyright enforcement).

276. *See supra* Section IV.B.

277. Bambauer, *supra* note 265, at 871.

278. *See supra* Section III.B; *see also* Bantam Books, Inc. v. Sullivan, 372 U.S. 58, 70 (1963) ("Any system of prior restraints of expression comes to this Court bearing a heavy presumption against its constitutional validity.").

279. Bambauer, *supra* note 265, at 885.

restraints often lack the safety valves that are necessary to prevent vagueness, over-breadth, or content discrimination, and they are executed under the radar of judicial review.[280]

Accordingly, prior restraint is more likely to be constitutionally permissible when "a government openly admits it blocks access to material, describes clearly what content it filters, targets prohibited information precisely, and arrives at decisions through accountable mechanisms of governance." [281] The proposed statute would do exactly that. Indeed, it would be adopted through a transparent legislative process, it would specify which type of illegal content could be targeted and by whom, and most importantly, it would ensure the removal/blocking injunction is subject to judicial review.[282]

The type of content that should be addressed by the proposed statute is illegal. The illegality of the content must be proven by clear and convincing evidence. [283] Only after a court renders its final decision about the illegality of the content, could it order platforms to remove it or to prevent users from accessing it.[284] In such cases, since "the expected damage in the absence of prior restraint is significant" and the government met a demanding standard, censorship should be deemed not only acceptable but also necessary. [285]

B. The Takings Clause

Under the Takings Clause of the Fifth Amendment of the U.S. Constitution, private property cannot be taken for "public use" without the payment of "just compensation."[286] This means that the government can take private property only if the taking is for public use and when it does so, it must provide just compensation to the owner.[287] The Supreme Court has interpreted the Takings Clause to include regulatory takings.[288] Accordingly, a particular regulation of the use of private property may require just compensation. [289] To determine whether just compensation is indeed required, the Court considers the economic impact of the regulation and its impact on the owner's reasonable investment-backed expectations.[290]

---

280. *Id.* at 886.
281. *Id.* at 873.
282. *See infra* Section VI.A.
283. Bambauer, *supra* note 265, at 935.
284. *See* Lemley & Volokh, *supra* note 249, at 175.
285. Bendor & Tamir, *supra* note 243, at 1164.
286. U.S. CONST. amend. V (stating "nor shall private property be taken for public use, without just compensation").
287. Kelo v. City of New London, 545 U.S. 469, 496 (2005).
288. *See* Pa. Coal Co. v. Mahon, 260 U.S. 393, 415 (1922).
289. Daniel A. Lyons, *Virtual Takings: The Coming Fifth Amendment Challenge to Net Neutrality Regulation*, 86 NOTRE DAME L. REV. 65, 90 (2011).
290. Penn Cent. Transp. Co. v. New York City, 438 U.S. 104, 124 (1978).

A possible argument may be that the proposed statute is essentially a government invasion of the platforms' services, where the government mandates the installation of technological measures on privately held infrastructure. Website blocking could even be viewed as a permanent invasion.[291] Indeed, some website blocking techniques, such as domain name system ("DNS") blocking, depend on the installation of a filtering device between the end user and the Internet.[292] A somewhat similar contention has been raised with respect to Net Neutrality. Specifically, it has been argued that Net Neutrality provides content providers with "an unlimited, continuous right of access to broadband providers' private property for free. This access allows them to physically invade broadband networks with their electronic signals and permanently occupy portions of network capacity, all without having to pay the network provider for access."[293] Nevertheless, since platforms probably have no legal right to illegal content, it seems unlikely that the proposed statute would be classified as a regulatory taking.

Nonetheless, obligating non-liable private platforms to remove or block illegal content clearly inflicts implementation costs on the designated platforms.[294] Therefore, it might be necessary to ensure that they are reimbursed for their role as law enforcers. Unlike the case of injunctions against non-liable third-party platforms rendered in civil lawsuits, the enforcement mechanism of the proposed statute is commenced by the government, not by a private plaintiff. Therefore, it might be necessary to provide public funding for the costs platforms would incur in establishing and applying the technological measures that are needed in order to remove or block the designated illegal content.[295] However, since it might sometimes be hard to predict these costs in advance, especially if the court allows platforms to elect between different technical methods of content blocking, it would probably be necessary to "establish a process whereby ISPs can apply for reimbursement if they are able to document such expenses."[296]

VI. BALANCING THE TRADEOFFS

The previous two Parts established a preliminary framework for shifting from liability-based speech regulation by platforms to

---

291. *See* Loretto v. Teleprompter Manhattan CATV Corp., 458 U.S. 419, 426 (1982) (holding that "a permanent physical occupation authorized by government is a taking without regard to the public interests that it may serve.").

292. *See* Perel, *Digital Remedies*, *supra* note 118, at 27.

293. Lyons, *supra* note 289, at 93.

294. *See supra* Section IV.C.

295. Bambauer, *supra* note 265, at 934.

296. *Id.* at 935.

enforcement-based speech regulation by platforms. Such a shift could minimize the proliferation of illegal content online, if two proposed fixes would be adopted: first, expanding the power of courts, through judicial interpretation, to enjoin non-liable third party platforms that enable the distribution of illegal content online, and second, enacting a new statute that would allow designated law enforcement agents to file removal requests with the court after proving by clear and convincing evidence that the designated content is illegal. While deeply rooted doctrines in constitutional law seriously challenge these two proposed fixes, the potential harms that may be caused by the spread of illegal content justify the attempt to overcome them.

In the last Part of this paper, I provide a concluding tradeoff map that balances the values democracies can promote by shifting to mandatory enforcement-based speech regulation by platforms and the values they may put at risk if this shift is not carefully and restrictedly designed. I further present some safety mechanisms that will minimize the risk to some of these values in order to ensure a balanced and effective regime of online content regulation. Mapping and balancing these tradeoffs are especially important for considering reforms based on legislative processes such as the second fix advanced in this paper; this is because tradeoffs ultimately produce "rules that involve protection for countervailing interests."[297]

A. Judicial Oversight versus Innovation

Shifting to mandatory enforcement-based regulation by platforms will take the public function of content adjudication out of the hands of private platforms and put it into the hands of objective judges. Indeed, when platforms remove content for law enforcement purposes — either in compliance with a court order or in response to a government warrant — they fulfill a public function.[298] When the government intervenes in such ways with platforms' content removal choices, it effectively launders state action through private actors.[299] Removing or blocking content for public law enforcement goals should therefore be subject to constitutional scrutiny.[300] However, while users may believe they enjoy the same constitutional protections when they speak through media platforms as they do in the proverbial

297. *Id.* at 898.

298. *See* Keller, *supra* note 12, at 1.

299. *Id.* at 3.

300. Wendy Seltzer, *supra* note 168, at 176 (2010) (explaining that the DMCA arguably replaces the need to file suit for copyright infringement in court; thus, it provides an alternative enforcement mechanism, which could be exposed to constitutional scrutiny).

town square, their online speech is governed almost exclusively by contract.[301]

As demonstrated in Part III, enforcement-based regulation by platforms is currently implemented on a voluntary and nontransparent basis. Because these enforcement practices are executed on private grounds, they escape judicial review. Platforms essentially assume the role of judges when they determine if a specific piece of content should be removed because it violates their internal policies about objectionable content.[302] Because platforms are private parties, the various interests implicated by speech regulation — including public safety, users' freedom of expression, and access to information — may only be represented to the extent that they are aligned with the platforms' business interests.

Indeed, platforms' decision making regarding content presentation reflects a complicated blend of private business considerations and public concerns.[303] As observed by Jack Balkin: "The infrastructure of free expression increasingly is merging with the infrastructure of speech regulation and the infrastructure of public and private surveillance."[304] The accusation that platforms not only allow objectionable speech online, but basically promote it by the logic of their systems demonstrates why we cannot count on platforms to make unchecked determinations about the legality of online speech.[305]

It is not only the way platforms decide whether to remove objectionable content that escapes legal scrutiny under the current regime of voluntary-based speech regulation by platforms. Indeed, also lacking is a check on the manner in which the government or other interested third parties influence platforms' content removal choices. For this precise reason, Bambauer has argued that open and direct "hard" government regulation of speech is more legitimate than "soft" regulation of speech, which relies on the government deploying tangentially related laws to limit online speech.[306]

Anecdotal evidence indicates that governmental actors abuse their power to censor speech for reasons that are unrelated to public safety. For instance, the Commonwealth of Kentucky sought to have 141

---

301. Fradette, *supra* note 28, at 948.

302. *See* Elkin-Koren & Perel, *Separation of Functions*, *supra* note 154, at 871; *see also* Niva Elkin-Koren, *Contesting Algorithms: Restoring the Public Interest in Content Filtering by Artificial Intelligence*, 7 BIG DATA & SOC'Y, 1, 2 (2020).

303. Elkin-Koren & Perel, *Separation of Functions*, *supra* note 154, at 872.

304. Jack Balkin, *Old School/New School Speech Regulation*, *supra* note 33, at 2297.

305. *See* Natasha Lomas, *Youtube Under Fire For Recommending Videos of Kids With Inappropriate Comments*, TECHCRUNCH (Feb. 18, 2019, 1:31 PM), https://techcrunch.com/2019/02/18/youtube-under-fire-for-recommending-videos-of-kids-with-inappropriate-comments/ [https://perma.cc/8V5X-JD9C] (detailing recent allegations against YouTube which claim that the site's recommendation algorithm "pushes users into a pedophilia 'wormhole'", essentially "facilitating and monetizing the sexual exploitation of children").

306. Bambauer, *supra* note 265, at 867.

domain names for gambling sites transferred to the state's control, claiming they posed a threat to its citizens.[307] Nevertheless, the real motive behind the government's action was political; Kentucky sought to protect earnings for offline gambling, which had ties to the Governor's political campaign.[308] In a different case, former Navy chaplain and Colorado Assembly candidate Gordon Klingenschmitt launched a campaign to use the DMCA to shut down the YouTube account of People for the American Way's Right Wing Watch project, which comments on the political views of candidates like Klingenschmitt using these candidates' own words.[309] The motives here had obviously nothing to do with copyright enforcement, but were instead a despicable attempt to silence political criticism. Such examples show how governmental actors, such as the Commonwealth of Kentucky, and third parties, such as Klingenschmitt, can pressure platforms to remove speech where they cannot legally do so.[310] The risk here is that "governmental goals may be disguised as objectives of private firms, driven by financial or competitive motives."[311] Because these requests to remove content are submitted directly to the platforms, bypassing judicial review, there is no guarantee that they are constitutional.[312]

The proposed fixes discussed in this paper would eliminate much of the mystery which currently surrounds the voluntary removal of illegal content by platforms. First, to the extent that a platform is enabling the dissemination of tortious content, and the injured individual fails to remove it himself, a court would be authorized to enjoin the platform and order it to remove the content. The platform would be given fair notice and an opportunity to be heard if it opposes the removal.[313] These procedural safeguards would ensure that the platform is included in the judicial process from its commencement, instead of being notified by the plaintiff about a relevant court order only after it is issued. As Eugene Volokh recently found, a substantial portion of court orders submitted to Google were what he characterizes as "either obviously forged or fraudulent or at least

---

307. Bob Pajich, *Kentucky Attempts to Seize Online Poker Domains: State Files Case to Stop Online Industry*, CARDPLAYER (Sept. 22, 2008), https://www.cardplayer.com/poker-news/5121-kentucky-attempts-to-seize-online-poker-domains [https://perma.cc/NA7K-2BTB].

308. Mike Masnick, *Kentucky's Gambling Domain Name Grab Sets a Terrible Precedent*, TECHDIRT (Oct. 10, 2008 9:48 AM), http://www.techdirt.com/articles/20081009/1142502506.shtml [https://perma.cc/TYM5-QHZN].

309. *Attempt to Silence the Political Speech at Right Wing Watch*, ELEC. FRONTIER FOUND. (Dec. 8, 2013), https://www.eff.org/takedowns/attempt-silence-political-speech-right-wing-watch [https://perma.cc/53UU-H8EP].

310. *See* Bambauer, *supra* note 265, at 899.

311. *Id.* at 901.

312. Fradette, *supra* note 28, at 973.

313. *See supra* Section IV.A.2.

highly suspicious cases."[314] If courts could directly enjoin platforms, the incentive to file fraudulent orders would probably decrease dramatically because the deterrence of being identified would increase as platforms would expect to receive injunctions that are specifically directed towards them. Second, with respect to illegal content, designated enforcement agents would be able to file removal requests with the court, and these would have to pass judicial review. Subjecting these two processes to judicial review would guarantee that there is a full, comprehensive, and unbiased consideration of the various interests involved.

Against this major gain of having a judicial check over speech regulation, it is important to acknowledge the possible impact on the development of new and efficient means of governing speech. In particular, because the fixes discussed in this paper aim to minimize the discretionary role of platforms in addressing illegal content, they may also discourage platforms from developing more accurate, efficient, and innovative means of addressing illegal content than the tools of the court.[315] This is especially true in relation to digital mechanisms that are directed to the constantly evolving online world, such as blocking orders. A judge may order a platform to implement existing means for removing or blocking illegal content, whereas a platform acting independently can benefit from a competitive advantage if it develops new and improved means to detect and remove such content. While courts could also allow platforms to address illegal content using whatever means they prefer, to the extent that the court would limit their discretion, their incentive to innovate could be diminished. While it is possible to argue that platforms could be incentivized to innovate so as to avoid possible judicial intervention, a counterargument would be that they would probably refrain from translating complex determinations about content legality into detection technology and leave this complicated task for the court.[316]

Nevertheless, this concern might be overstated. First, the proposed fixes only address illegal content. The regulation of other harmful and objectionable content by platforms remains unfettered. Since engaging in voluntary content moderation is probably essential for platforms to retain their popularity among their users and

---

314. Carolyn E. Schmit, *Shedding Light on Fraudulent Takedown Notices*, HARV. L. TODAY (Dec. 12, 2019), https://today.law.harvard.edu/shedding-light-on-fraudulent-takedown-notices/ [https://perma.cc/TC7K-2RFV].

315. *See generally* Perel, *Digital Remedies*, *supra* note 118.

316 . *See, e.g.*, Kurt Wagner, *Mark Zuckerberg Says He's 'Fundamentally Uncomfortable' Making Content Decisions for Facebook*, VOX (Mar. 22, 2018, 10:40 AM), https://www.vox.com/2018/3/22/17150772/mark-zuckerberg-facebook-content-policy-guidelines-hate-free-speech [https://perma.cc/U8P4-PD9X].

advertisers, it is unlikely that they will stop improving their content moderation capabilities. Second and related, courts can separate the adjudication of the request to enjoin a non-liable platform and the implementation of its removal order. The implementation stage should be principally controlled by the platforms' technological proficiency, but under the ongoing supervision of the court, because the platforms have the best knowledge regarding content moderation techniques.[317] To benefit from such ongoing oversight, it will be further necessary to ensure that the injunction is flexible, as discussed next.

B. The Rule of Law versus Flexibility

Another important benefit of the proposed shift to mandatory enforcement-based speech regulation relates to the preservation of the rule of law. Generally, the rule of law has long been interpreted as comprising two basic ideas: first, that individuals should be governed by law rather than by the arbitrary will of others; and, second, that no person is above the law.[318] The law must be clear, so people can develop reliable expectations and make autonomous choices accordingly. The proposed fixes advance the democratic notion of the rule of law because they establish transparent legal mechanisms for removing illegal content.

First, these mechanisms would be limited to content held illegal by a court. To determine illegality, the courts would follow statutory law and settled case law, which are both publicly available. Voluntary content regulation by platforms, on the other hand, follows internal, private policies that are largely non-transparent.[319] Users know very little about content moderation by platforms.[320] Second, the proposed mechanisms for removing illegal content are ordered by objective judges, and not by private, interested platforms. Third, these mechanisms enable dispute and reconsideration of the scope and breadth of the injunctions, which could further strengthen their compliance with the rule of law.

Nonetheless, committing to the rule of law may come at the cost of flexibility, which is especially important for adjudicating online content. Content moderation policies are often context specific and

---

317. *See generally* Perel, *Digital Remedies*, *supra* note 118.

318. ALBERT V. DICEY, INTRODUCTION TO THE STUDY OF THE LAW OF THE CONSTITUTION 189–90, 193 (10th ed. 1959).

319. *See* Fradette, *supra* note 28, at 973; *see also* Catherine Buni & Soraya Chemaly, *The Secret Rules of the Internet: The Murky History of Moderation, and How It's Shaping the Future of Free Speech*, VERGE (Apr. 13, 2016), https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-redditcensorship-free-speech [https://perma.cc/T8Z4-LWLL].

320. *See* Perel & Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, *supra* note 81, at 499–500.

time sensitive. Expressions change their meaning over time, so the laws addressing them must be flexible. Consider, for instance, the way Facebook has been wrongfully deleting posts containing the word "dyke."[321] The use of the word "dyke" may be hate speech when directed as an attack on someone; however, if one posted a photo of herself with #dyke, to denounce homophobia and reclaim the word, removing the content would mean restricting that person's ability to use that word in a self-referential, non-derogatory context.[322]

Flexibility is important not only in relation to the legal standards that discern legal content from illegal content, but also in relation to the technological measures that are used to implement content removals or blockings. Any application of structured technological solutions to address illegal content must be able to adjust to a rapidly changing technological environment.[323] For instance, "blocking access to pirate websites could be easily circumvented if users and content providers conceal their online conduct by using VPNs, proxy services and the like."[324] In other cases, pirate content may migrate from one location to another.[325] Sticking rigidly to predefined standards to secure the rule of law may therefore come at the price of efficiency and accuracy.

It is possible, however, to mitigate these concerns. In particular, as I discussed at length elsewhere, the court should be able to use different managerial devices to assure the orders it issues are sufficiently flexible to the changing circumstances.[326] These devices may include engaging in ex post revision, using the advice of technical experts, and imposing duration limitations.[327] So, for instance, if a court orders a platform or ISP to block specific designated domain names where infringing content resides, but the content subsequently migrates to new domains, the court would be able to adjust the order to include the new domains through ex post

---

321. *See generally Facebook: Stop Discriminating Against Lesbians*, CHANGE.ORG, https://www.change.org/p/facebook-stop-discriminating-against-lesbians [https://perma.cc/VBV9-GEEX]; Lisa A. Mallett & Liz Waterhouse, *Facebook Has a Problem With Dykes*, LISTENING 2 LESBIANS (June 24, 2017), https://listening2lesbians.com/2017/06/24/facebook-has-a-problem-with-dykes/ [https://perma.cc/C7AH-JUSL]; Kenny Sharpe, *Users Face Consequences as Facebook Struggles to Filter Hate Speech*, GLOBE AND MAIL (July 27, 2017), https://www.theglobeandmail.com/life/facebook-faces-pitfalls-in-quest-to-filter-hate-speech/article35819000/ [https://perma.cc/873D-DWXA].

322. *See* Annabel Thompson, *The Controversy Around Facebook Banning Lesbians from Using the Word 'Dyke,'* THINK PROGRESS (July 12, 2017, 3:58 PM), https://thinkprogress.org/is-facebook-banning-the-word-dyke-3720433451ed/ [https://perma.cc/4WRC-7VQU].

323. *See* Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance by Design*, 106 CAL. L. REV. 697, 739 (2018).

324. Perel, *Digital Remedies*, *supra* note 118, at 35.

325. *Id.*

326. *See generally id.* at 43–50.

327. *Id.*

revision.[328] Similarly, if the court orders a platform to use a specific blocking technique, but afterwards, a more accurate technique is developed, the court should be able to order the platform to replace the old technique with the new one.

C. Public Safety versus The Free Flow of Information

Perhaps the core justification for making the proposed shift in the way online speech is governed relates to public safety. Indeed, online speech can pose a serious risk "to the unwary and the innocent in terms of sorting out what is true and what is false, what is safe and what is dangerous to children and others, what is beneficial or neutral and what is devastatingly damaging, such as hate speech or injurious falsehoods." [329] Democracies must apply the principle of free expression in a way that does not endanger public safety. [330] As acknowledged over a century ago by Justice Holmes in *Abrams v. United States*, while the Court should protect expression, even that which is hated and feared, it should do so only up until the point that the speech "imminently threaten[s] immediate interference with the lawful and pressing purposes of the law."[331]

Online speech can and does often impose such threats. The 2016 "Pizzagate" incident is one famous example. After reading a fake news story about child sex slaves being held at a Comet Pizza — a restaurant in Washington, D.C. — under the direction of Hillary Clinton, a North Carolina man drove to the restaurant and fired a rifle inside.[332] Another example relates to the conviction of Harold "Hal" Turner, a blogger and occasional radio talk-show host, for inciting violence against judges in a blog post stating that they "deserved to be killed."[333] Turner's speech was aimed at persuading third parties to act violently on his behalf. His speech deserved censure because it magnified the risk of violence by unidentified third parties, who presumably shared his political views and prejudices.[334]

However, democracies should refrain from overprotecting public safety at the price of free expression. It is especially important to

---

328. *See id*. at 44.

329. Barry R. Schaller, *The First Amendment in the Digital Age*, 25 SACRED HEART U. REV. 60, 66 (2009).

330. *See generally* CONNIE HASSETT-WALKER, GUNS ON THE INTERNET (2019).

331. Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

332. *See* Eric Lipton, *Man Motivated by 'Pizzagate' Conspiracy Theory Arrested in Washington Gunfire*, N.Y. TIMES (Dec. 5, 2016), https://nyti.ms/2h4lMja [https://perma.cc/J2SC-R2BH].

333. Martha Neil, *Shock Jock Hal Turner Gets 33 Months for Threatening 7th Circuit Judges in Blog Post*, A.B.A. J. (Dec. 21, 2010, 8:38 PM), https://www.abajournal.com/news/article/shock_jock_hal_turner_gets_33_months_for_threatening_7th_circuit_judges_in_ [https://perma.cc/KN3S-CUDD].

334. *Id.*

ensure that the government does not over-predict violence from speech, seeking to suppress speech based on fear or dislike of radical ideas or speakers.[335] The balancing point must guarantee that legitimate speech remains freely accessible.[336]

Therefore, the proposed shift to enforcement-based speech regulation begins with addressing illegal speech. Such speech makes the easiest case for governmental intervention for it is "difficult to object to blocking access to material that users could not lawfully possess."[337] It is true that by limiting the proposed reform only to content that was adjudicated by clear and convincing evidence to be illegal, other harmful content, such as the fake news story which led to the Pizzagate shooting, will remain unaddressed. However, justifying the regulation of fake news is extremely controversial, so before making any reforms in the way such speech is treated, it is better to begin regulating where the potential risks imposed by the content are largely agreed upon.[338]

Another important safety mechanism in ensuring an appropriate balance between public safety and the free flow of information is technological. The proposed technological measures discussed in this paper must be employed carefully so as not to accidentally block protected speech. As an example, Pennsylvania's effort to block access to child pornography by requesting U.S.-based ISPs to prevent access to child pornography sites through cheaper blocking measures, such as IP blocking, resulted in the accidental censorship of numerous unrelated sites.[339] To safeguard against the targeting of protected speech as a side effect of blocking or removing illegal content, it is important to make sure platforms apply the most accurate means available. In particular, making these injunctions limited in duration and enabling ex post revision could minimize over-blocking.[340]

VII. CONCLUSION

Legislatures around the world are increasingly forcing prominent platforms, such as Facebook, Google, and Twitter, to rapidly remove

---

335. Lyrissa Barnett Lidsky, *Brandenburg and the United States' War on Incitement Abroad: Defending a Double Standard*, 37 WAKE FOREST L. REV. 1009, 1018 (2002).

336. Reno v. Am. C.L. Union, 521 U.S. 844, 865 (1996); *see also* Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 641 (1994) (holding that freedom of expression also implies limits on the government's ability to impede listeners who wish to hear that speech).

337. Bambauer, *supra* note 265, at 928.

338. *See generally* Evgeny Morozov, *Can The US Government Stem the Tide of 'Fake News' in a Postmodern World?*, GUARDIAN (Oct. 31, 2019, 6:00 AM), https://www.theguardian.com/global/commentisfree/2019/oct/31/can-the-us- government-stem-the-tide-of-fake-news-in-a-postmodern-world [https://perma.cc/ND3K-MWEB].

339. *See* Ctr. for Democracy & Tech. v. Pappert, 337 F. Supp. 2d 606, 633 (E.D. Pa. 2004).

340. *See generally* Perel, *Digital Remedies*, *supra* note 118, at 43–44, 46–47.

or block illegal content, including inciting materials, terrorist propaganda, and copyright infringement.[341] The United States, in contrast, holds firmly to the principle that the government cannot interfere with freedom of expression. The solid safe harbor established by § 230 of Communications Decency Act, which shields platforms from liability for harms caused by content they have not published, prevents most attempts to regulate unlawful speech. The result should worry us: necessary democratic means to protect public safety and secure individual rights in a world of digital communications are currently lacking.

Of course, the government and injured third parties could resort to less democratic ways to deputize non-liable platforms as law enforcers and directly request them to remove objectionable content. This form of soft censorship, however, could be more dangerous to freedom of expression. When illicit content is removed on the basis of non-transparent requests, which are adjudicated in accordance with privately developed removal guidelines, U.S. constitutional safeguards are put at serious risk. The rule of platforms replaces the rule of law; boilerplate terms of service replace balanced decision-making by lawmakers.

Speech regulation can no longer exclusively rely on platforms' voluntary, unchecked cooperation. Platforms have incredible enforcement capabilities, which should be harnessed efficiently and legitimately. Content found to be illegal must be removed from the Internet. Courts should be able to order non-liable platforms to use their most accurate and efficient means to remove such content.

After identifying what currently obstructs courts' ability to issue and enforce such orders, this paper proposed two necessary legal fixes. First, it proposed a legal interpretation that would allow injunctions against non-liable platforms in civil cases. Second, it recommended enacting a new, open, and transparent statutory legal procedure that would allow designated law enforcement agents to request the removal of content that was proven to be illegal by clear and convincing evidence. By addressing unlawful content through accountable channels, these fixes would fit well within the constitutional framework of freedom of expression and due process. By providing adequate notice to platforms, giving them an opportunity to be heard, and assuring they are reimbursed for the costs they incur when implementing removal orders, the proposed fixes will also secure due process.

In pursuit of a safe and lawful online environment, doctrines of platform liability often lead us to a deadlock. At the same time,

---

341. *See, e.g.*, *Facebook Must Delete Hate Postings, Austria Court Rules*, *supra* note 3; Alba, *supra* note 3; Meyer, *supra* note 3.

voluntary enforcement by platforms bypasses constitutional restraints. Regulating unlawful speech through scrutinized legal processes that would mandate the removal of such content by platforms is a rational solution to the problem of unlawful speech. Not only would it make the Internet safer, but it would also promote important democratic values, including the rule of law and accountability, which are lacking in the current, privately-run regime of speech regulation.