Platform for harm

Internet intermediary liability in Canadian law

September 2020

FRIENDS

Foreword

The early internet held out the promise of Eden but delivered anarchy instead. Nowhere is this anarchy and its consequences more evident than on social media, particularly Facebook and YouTube.

This timely and essential paper documents how these platforms have been overrun by hate speech, threats, and other illegal or otherwise objectionable content. It makes a compelling argument that Canadian law, while certainly imperfect, is decently equipped to beat back the online harms—individual, social, and political—with which platforms like Facebook are increasingly synonymous.

This paper reveals that the platforms' standard excuses for failing to address the scourge of illegal content—for example, claiming not to know about it until it is flagged by users—are likely insincere. Among the paper's many striking insights is that social media platforms—by their own admission—have technology that correctly interprets even highly nuanced user-generated content before they publish it. What is most shocking is not the ghastly prevalence of illegal content online, but rather how little governments have done to crack down on the businesses that profit from it.

These trenchant findings raise major questions about why the Canadian government has neglected to penalize platforms for their involvement in spreading harmful content. These companies do not just disseminate harmful content to billions of people, *they proactively recommend it*. Recently, *The Wall Street Journal* reported that almost two-thirds of extremism on Facebook is the result of the company's own recommendations. Yet, lawmakers have failed to properly consider—let alone address—the complicity of platforms in the myriad of harms for which they internally concede responsibility.

If we wish to remain a lawful, civil society, we must do better. Canada's leaders must answer the call and take meaningful steps to reduce online harms. This paper shows us how they can: by holding platforms to legally-binding standards of care, imposing penalties commensurate with the platforms' large revenues when they fall short of those standards, and ensuring that citizens can avail themselves of the weight of the state when pursuing complaints.

This last point is critical. This paper highlights that in the current system, the only way to penalize platforms for spreading illegal or otherwise objectionable speech is for individual complainants to take them to court. This allows platforms to exploit the titanic imbalance of power between themselves and their individual users by executing a, "war of attrition strategy that seeks to exhaust complainants' financial and emotional resources." Given these firms' deep pockets, it is no surprise that this strategy works. This also explains the relative lack of jurisprudence in this area—too few cases make it to trial.

If this precludes enough judges from weighing in, Parliament and prosecutors can intervene. Ultimately, this is about applying longstanding Canadian law to powerful companies that have so far profited from the ability to behave as though their self-imposed rules and standards supersede the laws and regulations of sovereign states like Canada.

The paper is clear that this is not about censoring political opinions or stifling free expression. Rather, it is about addressing the question of "platforms' liability for content that is already deemed inappropriate," such as child sexual abuse imagery or the incitement of violence. Who among us is against protecting our children from predation? Who among us believes that incitement to genocide is acceptable?

This paper also encourages us to consider how the prevalence of abuse and threats actually restricts freedom of expression by pushing some individuals—especially women, non-binary people, and people of colour—out of the public debate. The trolls may claim the mantle of heterodoxy, but ultimately, they reduce the diversity of expression online.

This paper is exceptionally thorough but also accessible. I strongly encourage Canadian elected officials, civil servants, judges, and lawyers to read it with care.

The implications could not be more stark. In barely two decades, platform giants have effectively unwound some of the foundational legal and social norms that allow people in democratic societies to live together peacefully. This should be cause for great concern.

This report is an urgent call to action for public and legal leaders: if you value democracy, justice, safety, and civility, you must exercise your power to hold those who threaten the public good responsible for the harms they inflict. Canadian leaders must reassert their powers as democratically elected representatives. The good news is that they do have such powers. The question now is whether they will continue to behave as though they don't.

Daniel Bernhard
Executive Director
Friends of Canadian Broadcasting

Executive Summary

This report explores the conditions under which Canadian law would hold internet intermediaries, such as social media platforms, liable for disseminating harmful content.* Online harms are diverse and widely evident on social media platforms, including hate speech, terrorism/radicalization, bullying, disinformation, encouragement of self-harm or suicide, among others.

Based on present principles of legal liability in Canadian common law, an internet intermediary that (1) actively promotes user content by way of algorithmic manipulation; and/or (2) receives notice from a prospective plaintiff of unlawful content, is very arguably a "publisher" at law; therefore liable for this content. This analysis is based on the law of "publication," but is not limited to traditional forms of publishing.

Large social media platforms often present themselves as passive parties to this content. However, certain platforms have demonstrated advanced capacities to understand users and user-generated content *before it is posted*. These capacities are used for the purpose of targeting users with revenue-generating content and advertisements, and for the purpose of removing content that platforms *themselves* deem to be "core" harms. Platforms have strong financial incentives to present users with content that retains their attention, and research in this field demonstrates that hateful content attracts more attention than moderate content. Given their sophisticated prepublication knowledge of content—including harmful content—it is reasonable to ask whether the prevalence of hateful content on such platforms is intentional.

Given these facts, Canadian law very arguably provides complainants with sufficient grounds to hold intermediaries liable for harms that take place on their platforms. However, the burden falls upon individual complainants to pursue these cases in court—alone. The significant imbalance of power and resources between that of platforms and individuals greatly reduces the prospect of a complainant having their case heard, which reveals a significant barrier to justice.

Online harms are clear and present, as are the financial incentives that deter internet intermediaries from taking meaningful action to end them. If Canadian leaders and policymakers wish to address the issue of online harm, they must intervene.

Summary of Recommendations to Canadian Policymakers

1. Protect freedom of expression and acknowledge its balance with that of other rights. The right to free expression must be upheld for all Canadians, as must protection against defamation, hate

* **Mark Donald**, legal consultant for the paper, is a lawyer specializing in defamation and privacy litigation. **George Carothers**, non-legal contributor for the paper, is Director of Research at FRIENDS of Canadian Broadcasting.

speech, and other harmful and illegal communications. Any regulation of internet intermediaries should retain and preserve this balance.

2. Acknowledge that certain online harms are categorically unacceptable. Incitement to suicide, stalking, and threats are emblematic examples. Responding robustly to such online harms is neither an alien nor a draconian principle in Canadian law.

3. Appreciate the platforms' technical prowess. Internet intermediaries possess advanced tools for targeting content and monitoring users. Regulators should acknowledge and engage these technological capacities.

4. Take a sovereign approach. Policymakers have developed Canadian rules for multinational firms in regulated sectors like telecommunications, finance, aerospace, and defence. Firms operating in the online sphere should be treated no differently.

5. Use enshrined legal principles as a starting point for appropriate regulation. Internet intermediaries appear to meet clear standards for liability that are present in Canadian common law.

6. Ensure the onus does not fall on individuals. Each harmed individual should not be burdened with initiating and financing a lawsuit against an internet intermediary. Consider the potential of a government agency, similar to the Privacy Commissioner's office, that could take investigative and enforcement actions on the behalf of citizens.

7. Apply meaningful and proportionate sanctions. Policies and penalties should proportionally fit the size of the company and the magnitude of the harm done.

8. Create and enforce strict privacy protections that minimize intermediaries' ability to collect personal data. Platforms benefit from the circulation of extreme content because it increases user engagement. This increased engagement generates vast amounts of user data that is captured in secret and used to direct content and advertisements at users. Reining in this invasive business model can reduce the circulation of harmful content.

9. Move promptly. Canada is uniquely placed to be a world leader in addressing online harms. By moving quickly, policymakers can shape standards that reflect Canada's legal regime, values, and priorities.

Setting the Scene: Internet Intermediaries and Online Harm

> "...[T]he most important decisions affecting the future of freedom of speech will not occur in constitutional law; they will be decisions about technological design, legislation and administrative regulations, the formation of new business models, and the collective activities of end-users."[1] — J. M. Balkin

In 2017, the British Columbia Supreme Court dealt with a case where the defendant falsely called her neighbour a pedophile on Facebook.[2] The neighbour was a teacher. Within 24 hours, the defamatory post spawned 57 other posts that falsely maligned the teacher, calling him "'pedo,' 'creeper,' 'nutter,' 'freak,' 'scumbag,' 'peeper,' and 'douchebag'" among other terms.[3] The court determined that the defendant was liable for defamation, not just for her own posts, but also for those made by her Facebook friends in response. Yet Facebook, which provided the platform for the defamatory remarks, recommended them to others, and provided a forum for the comments to be liked and shared, was not made a party in the proceeding. Without Facebook, this act of defamation may not have been possible. Yet, it bore no legal responsibility for the harm that transpired.

This paper explores the reasons why Canadian law has yet to fulsomely engage with questions of internet intermediary liability, and asks whether it is appropriate for Canada to take more robust steps to regulate internet platforms like Facebook in order to limit harms that are perpetrated on their platforms.

The internet is the greatest tool for informational exchange since the printing press, and Canadians have a clear interest in protecting freedom of expression. But the general lack of regulation, oversight and hierarchy in online publishing (i.e., the fact that anyone can publish anything anonymously from anywhere) also makes a laptop a dangerous tool in the hands of the wrong person. Over the past two decades, Canadian courts have come to recognize the manifold harms that arise from the internet, and particularly, the internet forums and platforms that allow for

1 Balkin, J. M. (2009) The Future of Free Expression in a Digital Age. Pepperdine Law Review. 36 (1). P. 427. See also, Balkin, J. M. (2004) Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society. New York University Law Review, 79 (1), P. 1-55.

2 *Pritchard v. Van Nes,* 2016 BCSC 686.

3 *Pritchard v. Van Nes*. Columbia Global Freedom of Expression. https://globalfreedomofexpression.columbia.edu/cases/pritchard-v-van-nes/ [Accessed May 27, 2020.]

instantaneous mass communication. In response, Canadian judges are making larger damage awards and more robust injunctive orders to halt online trolls— particularly in defamation cases.[4]

While the legal principles to hold individual persons responsible for harmful, defamatory, or illegal online behaviour are fairly developed (or at least, developing), Canadian law remains murkier when it comes to the appropriate level of liability for the internet platforms that are facilitating this illegal behaviour. Addressing this question is one of the next frontiers in Canadian legal and public policy thinking.

Some argue that the United States-Mexico-Canada Agreement (USMCA) prevents Canada from holding internet intermediaries liable for the user-generated content that they disseminate.[5] Indeed, the USMCA does appear to protect intermediaries from liability for user-generated content.[6] However, as argued herein, companies should not be indemnified from their basic obligations to society, particularly when their business models involve *recommending* content to users, including potentially harmful content. Furthermore, contrary to the USMCA, US lawmakers are already pursuing domestic legislation to qualify or reduce platforms' broad indemnity for user-generated content,[7] destabilizing the argument that the trade deal precludes Canada from acting.

The harms that take place on internet platforms are many and varied. Trolling, *doxing*, threats, sextortion, child sexual abuse imagery, bullying, industrial-scale propaganda, invective targeting of ethnic, social or religious minorities, incitement to suicide, junk science, and other common online harms are widely regarded as unacceptable in Canadian society.[8] However, Canada's existing laws and regulations were not designed to deal with the "weaponized virality" that social media can create.[9] Accordingly, Canadians should consider measures to address the mass transmission of such content by multibillion-dollar businesses that often have a strong financial incentive to

---

4 For an example, see the legal consultant's article, *Bringing order from chaos: some thoughts on recent judicial approaches to online libel cases,* The Advocates' Journal. Winter 2019. https://www.scribd.com/document/439971183/Bringing-Order-from-Chaos-Mark-Donald-The-Advocates-Journal-Winter-2019 [Accessed May 26, 2020.]

5 For an example, see Michael Geist's analysis *Why the USMCA Locks in the Internet Platform Liability System in the US, Canada and Mexico*. https://www.michaelgeist.ca/2020/05/why-the-usmca-locks-in-the-internet-platform-liability-system-in-the-u-s-canada-and-mexico/ [Accessed June 2, 2020.]

6 See section 19.17 of the *United States-Mexico-Canada Agreement*. https://ustr.gov/sites/default/files/files/agreements/FTA/USMCA/Text/19-Digital-Trade.pdf [Accessed June 4, 2020.]

7 For an example, see *Law Bites*. June 8, 2020. *Episode 54: Eric Goldman on Internet Platform Liability and the Trump Executive Order.* Michael Geist. https://www.michaelgeist.ca/podcast/episode-54-eric-goldman-on-internet-platform-liability-and-the-trump-executive-order/ [Accessed June 10, 2020.]

8 "Doxing" refers to the malicious act of making someone's private information publicly available on the internet

9 Bowers, J., Zittrain, J. (2020) *Answering Impossible Questions: Content Governance in an Age of Disinformation*, The Harvard Kennedy School (HKS) Misinformation Review, 1 (1). *[Age of Disinformation.]*

operate with minimal regulation or restriction. At the same time, Canadians must also consider the damaging restrictions on freedom of expression that increased regulations might inadvertently impose, and do everything possible to protect a free and open public discourse befitting of a healthy democracy.

This paper does not seek to debate exactly what constitutes "harm" online. Rather, it is concerned with what responsibility, if any, internet platforms should have for those harms that are already determined to be harmful or illegal, such as defamation, cyberbullying and other similar activities. This paper will therefore take the same approach to "harm" that American Supreme Court Justice Potter Stewart took to "obscenity" in a decision from 1964:

> "I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it. . ."[10]

Debates about tackling online harms engage the platforms who profit from them, and often hinge on the definition of who is and is not a "publisher" at law. Internet intermediaries, ranging from internet service providers to social media giants, are all involved in the act of publication in some sense or another, but the level of their involvement, and therefore responsibility, is not always intuitive. Many Canadians are likely to think that the publisher of harmful content is the one who posts the content, not the platform they post it to. What is more, the indeterminate definition of the term "publisher" in Canadian law means that the last decade or so has seen ongoing conflict in the courts over whether these internet intermediaries can, or even should, be considered publishers of the material that they host, transmit, and recommend to their users.

The answer to that question will have important and lasting consequences for how internet intermediaries are regulated, and how ordinary Canadians are able to protect themselves from potential harms, such as defamation, exploitation, bullying, and privacy breaches.

Roadmap for the Paper

This paper begins by summarizing the scope and extent of some of the main social harms that can

10 *Jacobellis v. Ohio*, 378 U.S. 184.

be facilitated by internet intermediaries: hate speech, terrorism/radicalization, proliferation of disinformation, cyber bullying, and the perpetuation of self-harm or suicide.

It then gives a summary of the relevant principles of traditional legal liability and suggests that using present Canadian common law principles, internet intermediaries who either (1) take active steps to promote content using algorithms or similar user-targeting systems; or (2) are given notice of unlawful content by prospective plaintiffs, are "publishers" at law and responsible for the content that they host.

The paper then summarizes the practical state of the law and argues that the avenues for legal redress available to ordinary Canadians are largely illusory, oftentimes due to the complexity and expense of legal proceedings, and from an imbalance between individual users and faceless multibillion-dollar, multinational firms.

We conclude with a general policy summary, discussing the strengths and weaknesses of the arguments coming from each side of the intermediary regulation debate. This leads into a discussion of the astounding technological prowess shown by some of these intermediaries in dealing with robust moderation of what they see as "core" harms to be moderated. With this prowess in mind, the question is then asked, what, if anything more, can or should intermediaries be required to do to fulfill their legal obligations to society.

To be clear, this paper argues for a robust investigation by policymakers into any and all policy tools which might help balance the positive effects of internet intermediaries with the harms they facilitate and perpetuate. While the paper does not make detailed policy prescriptions of its own, it is our hope that it will stimulate debate amongst Canadian policymakers as to how to properly strike that balance.

There are prominent voices in Canada and elsewhere who strenuously oppose the idea that companies like Facebook should be liable for publishing and recommending defamatory, hateful, or other illegal content. Almost invariably, their argument is that any such regulations would restrict freedom of expression to an unacceptable degree. We take the position that this approach is overly simplistic and does not strike the appropriate balance between user and platform, particularly in light of *existing* Canadian legal principles.

It is important that one point be made crystal clear: *nothing* in this paper should be interpreted to

suggest that the authors expect or welcome limitations on free expression and robust debate on issues of public interest. Rather, the goal is to highlight the social and public health issues that arise from internet intermediaries' business of spreading content widely and rapidly, including content that is accepted as harmful and/or illegal. This includes the very real problem of harmful content forcing certain people off of social media altogether, thereby reducing their opportunities for free expression and impoverishing public dialogue more broadly.

Our goal is to initiate a discussion on what, if any, regulatory responses are appropriate. After all, there is ample evidence that exposure to hatred, bullying, and harassment online dissuades many people from participating in public debate, particularly women, non-binary people, and racialized people.[11]

The line between objectionable expression and unlawful publication may be grey, blurred, or indeterminate, but it should not signal that a threshold for unlawful speech under Canadian law is so amorphous and indeterminate that it does not merit debate. As unsatisfactory as it may be, it must be acknowledged that even if its exact location cannot be determined, *a line between acceptable and unlawful speech does exist, and the general contours of this line can be made visible*. We all have a duty to attempt to further define its boundaries and to establish clear, fair rules for those people who cross it and the companies that help them do so.

11 *Online Harms White Paper*, Presented to Parliament by The Secretary of State for Digital, Culture, Media & Sport and the Secretary of State for the Home Department, Apr. 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf [Accessed May 26, 2020.][*Online Harms.*]

Coming to Grips with Online Harms

In the approximately two decades that the internet has come to dominate everyday life, online publishers have been subject to very little content regulation beyond their own internal policies and the extent to which litigants are willing or able to challenge the publisher's action or inaction in court. The largest intermediaries that Canadians engage with are American, and the American legislation under which they originated and grew is decidedly hands-off. In 1996, the United States introduced Section 230 of the *Communications Decency Act*, which explicitly exempts websites hosting user-generated content from liability for that content, while also exempting them from liability when they choose to remove content. The thinking at the time was that protecting intermediaries against legal action from users whose content was taken down would incentivize platforms to act in good faith and police the content they host as ethically and proficiently as possible.[12]

Platforms like Facebook, Google, YouTube, and Twitter originated under this regime, and accordingly, these firms strongly support Section 230. Indeed, for most of today's largest internet companies, the law happens to serve their interests particularly well. Many of these companies claim that their privileged position as pre-eminent facilitators of free expression and community building should take precedence over any harms that might result from the operation of their businesses. Speaking to the specific harm of misinformation, considerable amounts of which has circulated on his company's platform, Facebook CEO Mark Zuckerberg said, "I don't think most people want to live in a world where you can only post things that tech companies judge to be 100 per cent true."[13]

This legal regime has allowed companies like Facebook and Google to reach unprecedented levels of profitability. These firms now dominate the global advertising market, in part because they do not face the same liability risks that their competitors in traditional publishing and broadcasting do. In 2019, Google and Facebook—two companies—claimed as much as 60 per cent market share for digital advertising in the US.[14]

This relatively frictionless regulatory environment has enabled unprecedented amounts of content

12 Citron, D. K., & Franks, M. A. (2020) *The Internet as a Speech Machine and Other Myths Confounding Section 230 Speech Reform*. Boston University School of Law, Public Law & Legal Theory Paper No. 20-8.

13 Romm, T. (2019) *Zuckerberg: Standing For Voice and Free Expression*. The Washington Post. https://www.washingtonpost.com/technology/2019/10/17/zuckerberg-standing-voice-free-expression/ [Accessed May 26, 2020.]

14 Perin, N. (2019) *Facebook-Google Duopoly Won't Crack This Year*. eMarketer. https://www.emarketer.com/content/facebook-google-duopoly-won-t-crack-this-year [Accessed May 26, 2020.]

to flow through the major platforms; Facebook alone reports publishing 100 billion pieces of content *per day*.[15] This content covers everything from cat videos and baby pictures to propaganda, conspiracy theories, child sexual abuse imagery, terrorist content, incitement of genocide, and more. As discussed below, platforms often point to the sheer volume of content they publish as an argument for why it cannot be policed.

For the platforms, they claim that they are "too big to succeed" when it comes to content moderation. While the "too big to succeed" defence might sound reasonable on first blush, it remains prudent to examine whether companies like Facebook and YouTube lack the ability to reliably and accurately identify hateful or illegal content on their platforms.

Although the platforms' responsibility for online harms is up for debate, the harms themselves are clear and present dangers:

- In 2016, the Manitoba legislature passed the *Intimate Image Protection Act*[16] in order to assist victims with navigating the complexities of making legal claims or complaints against parties responsible for publishing and disseminating intimate images online without the victim's consent—a phenomenon sometimes colloquially referred to as "revenge pornography." Amongst other things, the Act creates a statutory framework for the tort of "non-consensual distribution of intimate images," as well as creating a broad framework for designated provincial authorities to assist victims with making and advancing complaints; facilitating the removal and destruction of impugned images, and if necessary, assisting victims with criminal complaints. As of mid-2018, 1,300 people had sought assistance under the legislation.[17]

- Globally, Facebook removed 8.7 million pieces of content in the third quarter of 2018 for breaching policies on child nudity and sexual exploitation.[18]

---

15 Wintour, P. (2020) *Mark Zuckerberg: Facebook must accept some state regulation*. The Guardian. https://www.theguardian.com/technology/2020/feb/15/mark-zuckerberg-facebook-must-accept-some-state-regulation [Accessed May 26, 2020.]

16 CCSM c. I87.

17 Kubinec, V. (2018) *More than 1,300 Manitobans seek help after intimate images shared*. CBC News. https://www.cbc.ca/news/canada/manitoba/revenge-porn-help-online-1.4637615 [Accessed May 26, 2020.]

18 Facebook (2018) Transparency Report. Available at: https://transparency.facebook.com/community-standards-enforcement#child-nudity-and-sexual-exploitation, cited in *Online Harms White Paper*, *supra*, at 1.6

- In 2018, the National Center for Missing and Exploited Children (NCMEC) received over 18.4 million referrals of child sexual abuse material from US tech companies.[19]

- According to the Internet Watch Foundation (IWF), roughly 55 per cent of child sexual abuse material found online depicts children aged ten or under, 33 per cent of this imagery is in the most serious category of abuse.[20]

- The FBI is now confronted with so much child sexual abuse imagery that they are forced to prioritize cases depicting infants and toddlers and are "essentially not able to respond to reports of anybody older than that."[21]

- All five 2017 terrorist attacks in the UK had an online element, and online terrorist content remains a feature of contemporary radicalization.[22]

- The 2019 Christchurch mosque massacre appears to have been designed for social media. The killer positioned the camera in such a way that resembles a first-person shooter video game and streamed live his rampage on Facebook to its global audience.[23]

- Facebook reported removing over 14 million pieces of content related to terrorism or violent extremism in 2018, the terrorist group Daesh (Isis) used over 100 platforms in 2018, making use of a wider range of more permissive and smaller platforms.[24]

- During the 2016 US presidential election, "20 top-performing false election stories" generated 18 per cent more engagements (likes, shares, etc.) than the "20 best-performing election stories from 19 major news websites."[25]

---

19 NCMEC. Available at: http://www.missingkids.com/footer/media/vnr/vnr2, cited in *Ibid*

20 Internet Watch Foundation (2017). Annual Report 2017. Available at: https://annualreport.iwf.org.uk/, cited in *Online Harms*, *supra*, at 1.7

21 *The Daily*, February 19 2020. *A Criminal Underworld of Child Abuse, Part 1.* The New York Times https://www.nytimes.com/2020/02/19/podcasts/the-daily/child-sex-abuse.html?showTranscript=1 [Accessed May 26, 2020.]

22 Speech at Digital Forum, San Francisco by the Rt. Hon. Amber Rudd, 13 February 2018, cited in *Online Harms*, *supra*, at 1.9

23 Marsh, J., Molholland, T. (2019) How the Christchurch terrorist attack was made for social media. CNN Business. https://www.cnn.com/2019/03/15/tech/christchurch-internet-radicalization-intl/index.html [Accessed May 26, 2020.]

24 Facebook (2018). Transparency Report. Available at: https://transparency.facebook.com/community-standards-enforcement#child-nudity-and-sexual-exploitation, cited in *Online Harms*, *supra*, at 1.10

25 Silverman, C. (2016) This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook. BuzzFeed News. https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook [Accessed May 26, 2020.]

As the examples above suggest, children and youth suffer a significant burden of online harms. A joint paper by the United Kingdom Home Office and Department for Digital Culture Media and Sport reports that nearly nine in ten UK adults and 99 per cent of 12- to 15-year-olds are online, with one in four of those users reporting experience of some form of harm related either to content or interactions online.[26]

As young people spend more time engaging with harmful content online, their mental health is deteriorating at an alarming rate. Researchers from the University of Toronto and the Hospital for Sick Children recently found that teenaged girls experienced a negative change in mood after spending just 10 minutes on Facebook, as well as a heightened desire to change their appearance and a greater likelihood of eating disorders.[27] They also note that in Ontario, the number of teenagers reporting moderate to serious mental distress increased by 63 per cent between 2013 and 2017, closely correlating to increases in smartphone usage and time spent on social media over that period. Between 2009 and 2014, the number of active monthly Facebook users grew by 480 per cent from 276 million to 1.3 billion.[28] Over that same period, the researchers found that "admissions to hospital for intentional self-harm increased by 110 per cent in Canadian girls."[29]

There is no indication that social media usage will decline in the future. After all, smartphones and the apps that operate on them are designed to be addictive.[30] Everything from the notification chimes to the flashing light indicating an unread notification are meticulously engineered by behavioural scientists whose talent is to transform users into modern versions of Pavlov's dog.[31] Tech executives often speak about their desire to create "habit-forming" products. Yet, unlike other harmful habit-forming products like tobacco and alcohol, there are no restrictions on youth using smartphones or social media, apart from the discretion of parents or guardians who are in many instances also addicted to their devices.

---

26 Ofcom and ICO (2018). Internet users' experience of harm online 2018. Available at: https://www.ofcom.org.uk/ research-and-data/ internet-and-on-demand-research/internet-use-and-attitudes/internet-users-experience-of-harm- online, cited in *Online Harms*, *supra*, at Joint Ministerial foreword and 1.3.

27 Abi-Jaoude, E. *et al.* (2020) Smartphones, social media use and youth mental health. *CMAJ* February 10, 2020 192 (6) E136-E141 https://www.cmaj.ca/content/cmaj/192/6/E136.full.pdf [Accessed May 26, 2020.]

28 Ortiz-Ospina, E. (2019) *The rise of social media*. Our World in Data. https://ourworldindata.org/rise-of-social-media

29 [Accessed May 26, 2020.] Abi-Jaoude *et al., supra*

30 *The Sunday Edition*, Sept. 14, 2018. *You can't stop checking your phone because Silicon Valley designed it that way*. CBC Radio. https://www.cbc.ca/radio/thesundayedition/the-sunday-edition-september-16-2018-1.4822353/you-can-t-stop-checking-your-phone-because-silicon-valley-designed-it-that-way-1.4822360 [Accessed May 26, 2020.] and Pringle, R. (2018) *Married to your phone? It's designed that way — but you can break up*. CBC News. https://www.cbc.ca/news/technology/cellphones-ramona-pringle-addiction-1.4637316 [Accessed May 26, 2020.]

31 Vengoechea, X., Eyal, N. (2015) *The Psychology Of Notifications*. Tech Crunch https://techcrunch.com/2015/02/05/the-psychology-of-notifications/

As with many health and social issues, online harms are not distributed evenly; women, racialized people, and other minorities often bear the brunt of the harm. For example, 21 per cent of women have received misogynistic abuse online. [32] Half of the girls aware of sexist abuse on social media say this has restricted what they do or aspire to in some way and/or forced some of them to leave social media. A study by the International Federation of Journalists found 64 per cent of female journalists had experienced death or rape threats, sexist comments, cyberstalking, account impersonation, and obscene messages, but almost half did not report it; over one-third had self-censored as a result.[33] In an analysis of some 70 million user comments left on its site over a decade, *The Guardian* found that of the ten most abused writers, eight were women and two were black men, even though the majority of the papers' opinion writers were white men.[34]

These examples show that it is not just people *per se* who are affected by online harms. Broader public ideals like a journalist's commitment to a story, women's sense of self-worth, and universal political enfranchisement are all threatened. If the unfettered circulation of hateful online speech discourages women, racialized communities, sexual minorities, or any other identifiable group from participating in public discourse, then the unregulated nature of social media and other online platforms can ultimately *suppress* free expression.

Online harms perpetrated on platforms like Facebook even include incitement to genocide. In 2018, a *Wired* magazine investigation concluded that without Facebook, a series of 2014 riots that swept across Myanmar would not have taken place. Despite multiple warnings to senior executives, the company was caught flat-footed when its platform started to circulate a false report that a Muslim shop owner had raped a Buddhist employee, triggering a wave of deadly violence that left tens of thousands dead and forced nearly a million more to flee.[35]

It is also important to consider the lasting effects of online harms—on the internet, nothing ever really disappears. Harmful publications and intimate personal details can be downloaded and remotely saved an infinite number of times, anywhere across the globe. Some European countries

---

32 Amnesty International (2017). The impact of online abuse against women. Available at: https://www.amnesty.org/en/latest/news/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women/, cited in *Online Harms, supra,* at 1.15.

33 *IFJ global survey shows massive impact of online abuse on women journalists*. International Federation of Journalists. https://www.ifj.org/media-centre/news/detail/article/ifj-global-survey-shows-massive-impact-of-online-abuse-on-women-journalists.html [Accessed May 26, 2020.]

34 The Guardian (2016). The dark side of Guardian comments. Available at: https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments, *Online Harms, supra,* at 1.25, Box 14.

35 McLaughlin, T. (2018) *How Facebook's Rise Fueled Chaos and Confusion in Myanmar*. Wired. https://www.wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar/ [Accessed May 26, 2020.] {*McLaughlin*]

have proclaimed "a right to be forgotten," which would require companies like Google to omit certain content from search results upon request.[36] Google, however, has resisted the demands of regulators to remove such content from search results globally, claiming that content should only be removed in search results of users who live in European jurisdictions. When the case was taken to the European Court of Justice, the court ruled in Google's favour, meaning that sensitive personal data of European citizens remains freely available to users who live anywhere outside of the court's jurisdiction.[37]

Canadian law has not previously contemplated such instant, perpetual and irreversible damage to an individual's reputation and well-being.[38] What is more, the phenomenon of perpetual availability threatens to eviscerate the presumptive "right to privacy" and "right to reputation" that Canadian law has attempted to uphold in principle for decades. [39] A recent Canadian study focusing on 15- to 21-year-olds' experiences of online harm highlighted a prevalence of vitriol against people identifying as Muslim, LGBTQ, and people living with disabilities.[40] Add to that the perception that social media platforms, for all their policies and behavioural guidelines, emphasize the importance of free expression but do not adequately address real-world harms:

> ". . .participants described a range of experiences with reporting harmful online content to social media platforms, involving everything from an Instagram page making fun of people with mental illnesses. . .to a racist comment on Instagram that suggested that the Obama children "look like gorillas." [41]

These Canadian youth seem to accept the platforms' inaction in the face of harmful content as normal or inevitable. Perhaps even more telling is that they don't expect it to change.[42]

Despite the clear and present harms that platforms facilitate, Canadian public policy and law has done little to engage with the social media companies that publish, promote, and recommend

---

36 Kelion, L. (2019) *Google wins landmark right to be forgotten case.* BBC News. https://www.bbc.com/news/technology-49808208 [Accessed May 26, 2020.]

37 Chee, F. *You have the right to be forgotten by Google - but only in Europe.* Reuters. https://www.reuters.com/article/us-eu-alphabet-privacy/you-have-the-right-to-be-forgotten-by-google-but-only-in-europe-idUSKBN1W90R5 [Accessed May 26, 2020.]

38 *Rutman v. Rabinowitz,* 2018 ONCA 80 at paras. 68-70.

39 *Jane Doe v. N.D.,* 2016 ONSC 541 at paras. 10, 39 (citing *Jones v. Tsige*, 2012 ONCA 32 at paras. 67 & 68), 57, varied on other grounds at 2016 ONSC 4920 [*N.D.*].

40 Bailey, J., Steeves, V. (2017) *Defamation Law in the Age of the Internet: Young People's Perspectives*. Law Commission of Ontario. June 2017. P. 38-39 [*Young People's Perspectives.*]

41 *Young People's Perspectives, supra,* at p. 50-51.

42 *Young People's Perspectives, supra,* at p. 53.

this content. While the platforms tend to adhere to the principles set forth in Section 230 of the US *Communications Decency Act*, we must remember that this is American law for which there is no Canadian equivalent. We must therefore ask what Canadian law has to say on these questions, and whether the US approach is compatible with our traditions, laws, jurisprudence, and values.

What is a Publisher?

If we wish to consider the platforms' liability for harmful content, we can begin by analyzing the facts through the prism of defamation law, or more specifically, the law of "publication." This field provides a body of jurisprudence that stretches back centuries and establishes helpful categories for determining when one actor is responsible for transmitting or, legally speaking, "publishing" the unlawful speech of another. The sorts of acts or omissions that make someone a publisher at law speak to the extent of their involvement in the published content.

Since "publication" in defamation law can serve as a shorthand for *prima facie* "legal liability" or "legal responsibility" in other areas of law like privacy, copyright, or criminal proceedings, defamation law's analysis of the question is most fulsome and most useful in conceptualizing the issue.

In Canada, the question of who or what a publisher is begins with an inquiry into just how involved the particular party is in the process of creating a statement or utterance:

> "Every participant in the publication incurs liability, regardless of the precise degree of his or her involvement. Primary participants, such as the writer, editor or newspaper company, are liable even for unintentional defamation not preventable by the exercise of due care. Persons who do not authorize publication but play the more subordinate role of mere distributors may escape liability on proof that they neither knew nor had reason to know or suspect that they were handling defamatory material."[43]

The legal question of "publication" is actually a continuum. On the one end, there exists the author—the person who directly makes the statement or utterance to a third party. On the other end is a party who is somehow *involved* with the publication, but in a way that is so passive—so bereft of knowledge of the content they are helping to circulate—that they cannot reasonably be deemed legally responsible for it. A prototypical example of such "innocent dissemination" is the postal worker who delivers a defamatory letter. The letter would not have reached the target without their participation, but to say that they are complicit in the defamation itself is a stretch too far.

Judges and legal academics have spilled significant ink debating just how to deal with the expansive

43 *«The Canadian Encyclopedic Digest*: Defamation; V; 2 – What Constitutes Publication at para. 111.

middle of this publication continuum. The general principle is that for a party to be a true publisher, they must have direct knowledge of the content being published. So, the printer that is employed to create physical copies of a defamatory newsletter is likely not liable as a publisher, but a newspaper owner who allows defamatory comments into print is liable. While any number of the publishing company's board or workers may not have specific knowledge of an unlawful publication, someone in the company approves all printed content and their *approval,* explicit or implicit, is grounds for liability. In the end, the question of publication is at its core one of control and/or knowledge; there are different standards for parties who know (or should know) what they are helping to circulate and those who do not. The more intimate or direct the involvement, the greater the risk of liability.

The crucial point is that this knowledge need not be advance knowledge. Canadian law tends towards the proposition that a party *becomes* a publisher when they are provided with notice of a defamatory publication that has already occurred.

The seminal case in this area of law is *Byrne v. Deane,* a 1937 ruling from the United Kingdom that subsequently made its way into Canadian jurisprudence. The trial dealt with a situation where persons unknown posted a defamatory message on a golf course bulletin board. The golf course's management was found to be liable for this defamation because they were informed of the posting and failed to remove it.[44]

The *Byrne* precedent asserts that publication is not just a deliberate act—it can also be a deliberate failure to act. Failing to remove harmful content after being notified of its existence therefore constitutes approval, adoption, promotion, or ratification of the defamatory material in Canadian law. If the libel is not taken down, then publication and responsibility are established.[45]

44 *Byrne v Deane*, [1937] 2 All ER 204 [*Byrne*] as cited by, for example the Supreme Court of Canada in *Crookes v. Wikimedia Foundation Inc.*, 2011 SCC 47 (S.C.C.) at para. 87 [*Crookes.*].

45 *Crookes, supra,* at paras. 97-98.

Are Internet Intermediaries Liable as Publishers Under Canadian Law?

Based on the *Byrne* standard, platforms like Facebook and YouTube would be legally liable for harmful and illegal content on their platforms if they (a) know about it in advance and decide to publish anyway, or (b) are notified of the offending content and fail to take it down.

On first blush, the latter principle is arguably established; all major platforms have a complaints process of some sort. Once a complaint is received, the platform becomes a publisher of the offending content under Canadian law and they are legally liable if they neglect to remove it.

However, as discussed below, it must be acknowledged that the *Byrne* principle has *not* been specifically applied to any major internet platform in case law. The principle therefore hangs as a Sword of Damocles over internet platforms in Canada without actually striking a blow. It is this practical gap which the authors submit could be addressed through robust public policy intervention.

Still, a more fundamental question arises—one which may make application of the *Byrne* principle redundant: do platforms have *advance* knowledge of the content they publish, and if so, does the subsequent act of publication constitute express approval?

In order to answer this question, we must examine how platforms actually work. Prior research and judicial/public policy silence on the issue reflects a poor understanding of the facts, and creates a significant blind spot in conceptual thinking over the appropriate balance between regulation and free expression.

How Internet Intermediaries Work (And the Harms that can be Facilitated)

In a 2017 review of the law surrounding online defamation, the Law Commission of Ontario identified three categories of internet intermediaries: (1) internet service providers (ISPs such as Bell, Rogers, Shaw, etc.) that send content to and from our devices; (2) search engines that respond to user-initiated queries; and (3) content hosts that operate message boards, and allow members to exchange views.[46]

46 *Defamation in the Internet Age: Consultation Paper*. Law Commission of Ontario (Toronto: November 2017) at page 34-35 http://www.lco-cdo.org/wp-content/uploads/2017/12/Defamation-Consultation-Paper-Eng.pdf [Accessed May 25, 2020.] [*LCO*]

On first blush, firms such as Facebook and YouTube qualify as "content hosts," but in fact they are more than just glorified message boards. They may look and smell a lot like the golf course notice board seen in *Byrne,* but the reality is altogether more complex.

Content curation is the central competency of Facebook, YouTube, Twitter, and other successful internet intermediaries. Recent research suggests that a majority of internet users— and presumably a majority of judges, legislators, and civil servants—are not fully aware that much of what they see online is meticulously personalized; chosen just for them by algorithms which implement the platforms' editorial policies.[47] A user's Facebook feed is not a complete, chronological repository of everything the user's friends post. Rather, the decision about what to include and what to exclude on the feed is made "in secret" by algorithms based on extensive stores of data harvested from a user's browsing history, social media activity, location history, wearable technologies, and "smart home" devices, among other sources of data capture. Most of the time this data is taken without the user's knowledge or consent, and users are not paid for it.[48]

Due to this process of curation, no two Facebook users will see the same content in their feed. Likewise, no two users watching the same YouTube video will see the same recommendations about what video to watch next.

Platforms' decisions about what to present and what to conceal is made with one objective in mind: retaining the user's attention for as long as possible. Successful platforms are exceptionally good at retaining one's attention. The process has six core steps:

1. Take as much personal data as possible, often surreptitiously or under threat of exclusion from popular services.[49]

2. Profile each user based on their data and behaviour to accurately predict their interests, mood, desires, and state of mind, with the goal of predicting what they are most likely to look at and what they are most susceptible to buy at any given moment.[50]

---

47 *Online Harms, supra,* at Harm: Online disinformation – Box 12.

48 In her seminal work on the mechanics of data monetization businesses, Shoshana Zuboff refers to this process as "the dispossession cycle." See Zuboff, S. (2019) *The age of surveillance capitalism: the fight for a human future at the frontier of power.* New York: Public Affairs. [*The age of surveillance capitalism*]

49 Zuboff calls this "the extraction imperative." See *The age of surveillance capitalism, supra*.

50 Zuboff calls the process "rendition," wherein people's lives (movements, relationships, profession, interests, purchases, hobbies, genetic material, etc.) are disembodied and "rendered" as data points. See *The age of surveillance capitalism, supra.*

3. Interpret all available content in the system to identify posts, links, videos, and other content that the user is most likely to look at given their profile and current circumstances (mood, location, time of day, time of month, etc.).

4. Present this content to the user to retain their attention.[51]

5. Exploit the user's attention to subject them to ads, each individually selected to maximize probability of a click, based on the user's profile and current circumstances (mood, location, time of day, time of month, etc.).

6. Intervene in (imperceptible and undisclosed) ways to increase the likelihood that a user clicks on an ad or discloses more personal data.[52]

Steps two through five cannot be accomplished without detailed knowledge of user-generated content. Platform companies' own statements and reports boast about their intimate knowledge of the content users provide. Facebook, for example, is able to inventory and categorize "hot" and trending topics worldwide, including topics of conversation, and behavioural shifts.[53] The company presents a yearly report on their findings as an enticement to advertisers.[54]

Google has similar capabilities, which it describes in the corporate language of "insight mining" —touting its ability to monitor and curate user interactions in order to help create and implement advertising and promotion strategies. Google tells advertisers how it monitors *"... search results in order to learn when people were talking, and what they were talking about*, and when they joined the conversation" [emphasis added].[55]

These firms are heavy investors in machine learning and artificial intelligence (AI) focused on

---

51 This is the interface users interact with when using popular services like Facebook. As one engages with the interface, they generate what Zuboff calls "data exhaust" or "behavioural surplus" which the platforms appropriate for their own use. See *The age of surveillance capitalism, supra.* And, for a fascinating contemplation of the concept of "the interface" and its elusive promise of fulfillment, see Kingwell, M. (2019). *Wish I were here: boredom and the interface.* Kingston: McGill-Queen's University Press

52 Zuboff refers to this as "the actuation imperative." See *The age of surveillance capitalism, supra.*

53 See, for example: *Hot topics in Canada for December 2019*. Facebook for Business. https://www.facebook.com/business/news/insights/2019-12-hot-topics-canada; and also, *Topics to watch in the United States for January 2020*. Facebook for Business. https://www.facebook.com/business/news/insights/2020-01-topics-to-watch-united-states [Accessed May 26, 2020.]

54 *The 2020 Topics & Trends Report from Facebook IQ*. Facebook for Business. Report downloaded from https://www.facebook.com/business/news/insights/2020-topics-and-trends-report [Accessed June 1, 2020.]

55 *Data and Insight Tools*. Think with Google. https://www.thinkwithgoogle.com/features/youtube-playbook/topic/data-insight-tools/ [Accessed June 1, 2020.]

comprehension.[56] This technology has resulted in the "virtual assistants" that now ship with every smartphone and are increasingly present in domestic settings, in the form of "smart home" devices like Google Home or Amazon Echo. These devices run software with human names like *Alexa* or *Siri* that have a near-human ability to comprehend and converse. In 2018, Alphabet CEO Sundar Pichai demonstrated the Google Assistant's remarkable conversational capacity as it called a hair salon to book an appointment, and was so realistic that the salon receptionist was unaware that she was talking to a computer.[57]

In 2017, Facebook rolled out robust AI tools that are able to monitor live video broadcasts and identify content that suggests suicidal ideation in real-time.[58] As early as 2013, Facebook conducted experiments on users—without their knowledge or consent—to test methods that succeeded in not only determining but also *manipulating* a user's emotional state. Facebook's researchers also found evidence of a "social contagion" effect, whereby changes in one user's mood were found to impact their contacts' moods in predictable ways.[59]

Such evidence suggests that viral publication of extreme content on these platforms is beneficial and perhaps intentional, rather than a bug or an accident. Wael Ghonim, Harvard scholar, former Google employee, and a leader of the 2011 Tahrir Square protests in Egypt, argues that platforms have clear financial incentives to present, "content that is gross, violent, or sexual [and] gossip which is humiliating, embarrassing, or offensive," because such content is proven to increase engagement. [60] Indeed, when it comes to increasing engagement, the more outrageous the content, the more effective it is. Ghonim cites a 2017 study by the Pew Research Centre, which found that "U.S. members of Congress received 50 percent more 'likes,' three times as many comments, and twice the number of shares for posts that expressed 'indignant disagreement' than for those that expressed bipartisan sentiments."[61] It is reasonable to assume that platforms, some of the world's foremost experts in machine learning and data analytics, have likely reached the same conclusions as Ghonim. It would therefore be prudent to examine whether the prevalence of hateful, inciteful,

---

56 See, for example, Gershgorn, D. (2019) *How Google aims to dominate artificial intelligence.* Popular Science. https://www.popsci.com/google-ai/ [Accessed June 1, 2020.]

57 Welch, C. (2018) *Google just gave a stunning demo of Assistant making an actual phone call.* The Verge. https://www.theverge.com/2018/5/8/17332070/google-assistant-makes-phone-call-demo-duplex-io-2018 [Accessed May 29, 2020.]

58 Callison-Burch, V. et al. (2017) *Building a Safer Community with New Suicide Prevention Tools*. Facebook. https://about.fb.com/news/2017/03/building-a-safer-community-with-new-suicide-prevention-tools/ [Accessed June 1, 2020.]

59 *The age of surveillance capitalism, supra,* at p.299-301.

60 Ghonim, W. (2018) *Transparency: What's Gone Wrong with Social Media and What Can We Do About It?* Harvard Kennedy School Shorenstein Center on Media, Politics, and Public Policy. https://shorensteincenter.org/transparency-social-media-wael-ghonim/ [Accessed May June 1, 2020.]

61 *Partisan Conflict and Congressional Outreach*. Pew Research Center. https://www.people-press.org/2017/02/23/partisan-conflict-and-congressional-outreach/, cited in *Ibid.*

aggressive, defamatory, or other objectionable content is deliberate. If hate drives engagement and engagement drives profit, a rational, profit-seeking actor could use its advanced capabilities to comprehend the substance of posts, videos, and photos, to push out as much engagement-generating content as possible.

Writing in *The New Republic,* Bob Moser argues that this is precisely what takes place on YouTube, calling the video sharing platform, "the worldwide leader in white supremacy."[62] The radicalization process is predictably gradual. The *Atlantic*'s Conor Friedersdorf argues that YouTube leads viewers down a "rabbit hole of extremism, while Google racks up the ad sales," because their algorithms promote fringe views.[63] Friedersdorf quotes scholar Zeynep Tufekci's observations of how YouTube recommends videos:

> "Videos about vegetarianism led to videos about veganism. Videos about jogging led to videos about running ultramarathons. It seems as if you are never "hard core" enough for YouTube's recommendation algorithm.
>
> It promotes, recommends, and disseminates videos in a manner that appears to constantly up the stakes. Given its billion or so users, YouTube may be one of the most powerful radicalizing instruments of the 21st century."[64]

This concept of "radicalization creep" is confirmed by former Google employees (Google owns YouTube through its holding company, Alphabet). Some engineers, such as Guillaume Chaslot, grew alarmed that the recommendation engines they built were increasing users' viewing time by pushing them to increasingly extreme conspiracy videos, misinformation, and hate. When Chaslot tried to modify the algorithm to produce more balanced recommendations, Google responded by terminating his employment.[65]

The implications for hate and radicalization are deeply concerning. A recent study by *Data & Society* documents how YouTube leads viewers from "mainstream versions of libertarianism and

---

62 Moser, B. (2017) *How YouTube Became the Worldwide Leader in White Supremacy*. The New Republic. https://newrepublic.com/article/144141/youtube-became-worldwide-leader-white-supremacy [Accessed June 1, 2020.]

63 Friedersdorf, C. *YouTube Extremism and the Long Tail*. The Atlantic. https://www.theatlantic.com/politics/archive/2018/03/youtube-extremism-and-the-long-tail/555350/ Accessed June 1, 2020.]

64 Tufekci, Z. (2018) *YouTube, the Great Radicalizer*. The New York Times. https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html [Accessed June 1, 2020.], cited in *Ibid*

65 *The Daily*. April 17 2020. *Introducing 'Rabbit Hole'*. The New York Times. https://www.nytimes.com/2020/04/17/podcasts/the-daily/rabbit-hole.html [Accessed June 4, 2020.]

conservatism, all the way to overt white nationalism."[66] There is now widely available evidence of politicians exploiting this frenzy of indignation for their own gain, and companies like Facebook have capitalized on this by allowing political leaders to not only target susceptible voters, but also to buy advertising, which Facebook itself deems to be false. Mark Zuckerberg has stated that, "people should be able to judge for themselves" whether a politician is truthful.[67] Senator Elizabeth Warren responded by calling Facebook a "disinformation-for-profit machine."[68]

Importantly, none of this is to infer that *all* internet intermediaries are alike. Indeed, some may have so passive a role in the publication process that they may be deemed either: (1) not to be legally publishers at all; or (2) in the alternative, be "publishers" who are able to easily avail themselves of the legal defence of "innocent dissemination."[69] However, many of the biggest, most influential internet intermediaries, the likes of Facebook, Google, YouTube, and Twitter, cannot credibly plead "innocent dissemination." The fundamental facts of their business, coupled with their own pronouncements, portray a different story.

The platforms' foreknowledge of the content they circulate is so detailed and profound that they very arguably are publishers under Canadian law.

Applying the *Byrne* principle to these self-moderating and algorithm-employing internet intermediaries, mining of users' personal information and the subsequent ranking and promoting of different content for financial gain certainly appears like approval, adoption, promotion, and ratification of content as discussed in *Byrne*—which is to say "publication." Whether one uses the more passive Byrne conception of publication, or the more activist "knowing involvement" standard favoured by the British, it can be argued that an algorithm-sporting internet intermediary would be a publisher under both benchmarks.

---

66 Lewis, R. (2018) *Alternative Influence: Broadcasting the Reactionary Right on YouTube.* Data & Society. https://datasociety.net/library/alternative-influence/ [Accessed June 1, 2020.]

67 Budryk, Z. (2019) *Zuckerberg on allowing political ads: 'People should be able to judge for themselves'.* The Hill. https://thehill.com/policy/technology/472571-zuckerberg-on-political-ads-people-should-be-able-to-judge-for-themselves [Accessed May 26, 2020.]

68 Kang, C., Kaplan, T. (2019) *Warren Dares Facebook with Intentionally False Political Ad*. The New York Times. https://www.nytimes.com/2019/10/12/technology/elizabeth-warren-facebook-ad.html [Accessed June 1, 2020.]

69 See Laidlaw, E., Young, H. (2017) *Internet Intermediary Liability in Defamation: Proposals for Statutory Reform*. Osgoode Hall Law Journal. 56 (1) P. 112-120

Due to the aggressive use of recommendations to retain user attention, intermediaries like Facebook and YouTube are arguably *more* "involved" in publishing users' content than the users themselves. Bowers and Zittrain capture the situation suitably:

> "When a library patron asks for *Mein Kampf* at the circulation desk, we expect, in large part as a matter of expressive freedom, that librarian will return with that hateful text in hand or direct the patron to a place where it can be found. At the same time, we would likely hope that an open-ended request for "something interesting to read" might turn up a different book."[70]

Yet, that is precisely how YouTube's recommendation engine works. It is now believed that as many as 80 per cent of YouTube's views are driven by recommendations.[71] Platforms are not necessarily responding to queries for harmful content. Rather, they are *recommending* this content to users, introducing them to it in order to increase engagement, retain their attention and make them available to advertisers.

70 *Age of Disinformation, supra*

71 See, for example, *After Truth: Disinformation and the Cost of Fake News (documentary film)* https://www.crave.ca/en/search?q=after%20truth#/special/47072 [Accessed May 29, 2020.]

The Current State of the Law

Explaining the Dearth of Judicial Guidance on Internet Intermediary Liability

By now, it should be clear that intermediaries like Facebook and YouTube are very arguably publishers according to Canadian law. Following the *Byrne* principle, they become publishers when they are notified of harmful content. But we have also seen compelling evidence that they have detailed knowledge of harmful, defamatory content before it is even published. That they publish it anyway suggests express approval, and therefore, complicity.

Yet despite the prevalence of this harmful content, Canadian law gives little guidance as to when intermediaries are liable for this content and to what extent. As such, the platforms have not been required to address their own involvement in the publication of harmful or illegal content.

While there is no shortage of case law that *involves* internet intermediaries in Canada in one way or another, the question of their liability for torts like defamation and invasion of privacy has never been truly tested in any fulsome sense. Nor is there any statute that would purport to explicitly attach such liability.

The video of the 2019 Christchurch mosque attack is a useful example. Facebook was alerted to the video's existence within 30 minutes, but despite a herculean effort to remove the video, the company admitted that 1.5 million copies were uploaded to its servers, and 300,000 of those uploads were then promoted to its users—all within 24 hours of the atrocity.[72] YouTube faced a similar situation.[73] So far, and to the best of the authors' knowledge, neither platform has faced any legal consequences in Canada for failing to remove this content after having been notified.

Likewise, Facebook has faced no consequences for allowing advertisers to illegally target advertisements based on race, age, gender, or other protected grounds.[74] After being flagged by journalists, and therefore notified, the company announced changes in policy. However, these changes would apply only in the United States, leaving illegal, discriminatory ad targeting to

72 Naughton, J. (2019) *Christchurch shows how social media sites help spread the poison of far-right ideology.* The Guardian. https://www.theguardian.com/commentisfree/2019/mar/24/christchurch-shows-how-social-media-sites-help-spread-poison-far-right-ideology-youtube-facebook [Accessed May 26, 2020.]

73 *Ibid*

74 McIntyre, C. *Facebook's ad system seems to discriminate by race and gender.* The Logic https://thelogic.co/briefing/facebooks-ad-system-seems-to-discriminate-by-race-and-gender/?gift=65be5b57c596fba32d6faf855124b577 [Accessed June 4, 2020.]

continue in Canada.[75] Investigations by *The Logic* revealed that it was possible to purchase housing ads that would explicitly exclude Indigenous peoples. While they were able to discriminate against Indigenous peoples in Canada, they were unable to discriminate against identifiable marginalized groups in the United States: terms like "'African American,' 'Black,' 'Mexican,' 'American Indian,' 'Italian,' 'queer' and 'transgender'" were off-limits.[76]

What is clear is that the principles set out in the *Byrne* decision cast a long shadow over lawyers' thinking and make it relatively simple for a plaintiff to effectively argue, at least in the opening salvos of litigation, that any intermediary defendant is arguably a "publisher" once they are given notice of allegedly defamatory content.[77]

But having an arguable case and proving one are two very different things. The legal consultant to this paper has, at times, seen counsel for intermediaries take hardline positions to support their refusals to remove content, until such time in the litigation when considerable money must be spent. In such cases, the legal consultant presumes that the intermediaries' strategy is to push on with litigation in the hope that plaintiffs might give up or simply run out of money before the real fight starts. In other cases, the legal consultant has seen internet intermediaries agree to remove allegedly unlawful material summarily, but only on the condition that plaintiffs agree to aggressively prosecute actions against the original authors of the content. The practical results are first, that average citizens—ones bearing the brunt of these online harms—are often "priced out" of accessible justice; and secondly, that the asymmetry of financial power between users and internet platforms acts as an impediment to seeing important cases go before the court, and therefore impedes clear judicial pronouncements on the issue of intermediary liability.

However, we have also demonstrated that intermediaries like Facebook and YouTube have detailed knowledge of defamatory and hateful publication in advance, and that whatever content is published is therefore published with their prior knowledge and authorization.

75 *Ibid*

76 McIntyre, C. (2018) Facebook allowed housing and employment advertisers to exclude users with Indigenous interests https://thelogic.co/news/big-tech/facebook-allowed-housing-and-employment-advertisers-to-exclude-users-with-indigenous-interests/?gift=6df096977caadabadf5c6517cbf8083e

77 *Pritchard v. Van Ness,* 2016 BCSC 686.

Policies, Principles and Best Practices – the Intermediaries Police Themselves

In the absence of clear legislation, regulation or jurisprudence, intermediaries have become *de facto* regulators and moderators of the content they host, accountable only to themselves. To name a few for purposes of example only, Facebook, Instagram, and Google all have broad terms of service/ user guidelines that permit them to monitor and police objectionable content.[78] These efforts at least demonstrate acknowledgement of some kind of responsibility, and an awareness of their own *possible* liability.

Given the lack of governmental movement on this issue, social pressure and the platforms' voluntary actions appear to be the driving forces behind increased moderation efforts. As far back as 2015, former Twitter CEO Dick Costolo admitted:

> "We suck at dealing with abuse and trolls on the platform and we've sucked at it for years. It's no secret and the rest of the world talks about it every day. We lose core user after core user by not addressing simple trolling issues that they face every day. . .I'm frankly ashamed of how poorly we've dealt with this issue during my tenure as CEO. It's absurd. There's no excuse for it. I take full responsibility for not being more aggressive on this front. . ." [79]

In the past few years, Facebook announced a shift to a *proactive* model that is able to remove 99 per cent of terrorist content and 80 per cent of hate speech from the platform before a complaint is made.[80] As early as 2017, Facebook reported that in some cases it was so proactive that it was able to filter out the vast majority of terrorist content before it was ever seen by a user.[81] This appears to be an admission that the platform *is* able to comprehend content posted to its platform, exercise editorial judgement and parse the harmful from the innocuous before publishing users' posts.

In his 2018 testimony before the US Senate's Judiciary Committee on data privacy and Russian

---

78 See, for example, Facebook's Community Standards: https://www.facebook.com/communitystandards/; Instagram's Community Guidelines: https://help.instagram.com/477434105621119; and Google's Terms of Service: https://policies.google.com/terms?hl=en-US.

79 Tiku, N., Newton, C. (2015) *Twitter CEO: 'We Suck at dealing with abuse'.* The Verge. http://www.theverge.com/2015/2/4/7982099/twitter-ceo-sent-memo-taking-personal-responsibility-for-the [Accessed June 1, 2020.]

80 See, for example, *Community Standards Enforcement Report, November 2019* Edition. Facebook. https://about.fb.com/news/2019/11/community-standards-enforcement-report-nov-2019/ [Accessed June 4, 2020.]

81 *Facebook's AI wipes terrorism-related posts*. November 29, 2017. BBC News. https://www.bbc.com/news/technology-42158045 [Accessed May 26, 2020.]

disinformation, Facebook CEO Mark Zuckerberg candidly acknowledged Facebook's growing responsibility to tackle the social harms taking place on its platform:

> "That goes for fake news, foreign interference in elections, and hate speech. . . "We didn't take a broad enough view of our responsibility". . . "It's not enough to just connect people. We have to make sure those connections are positive". . . "It's not enough to just give people a voice, we have to make sure that that voice isn't used to harm other people or spread misinformation". . ."I'm committed to getting this right."[82]

Certainly, public pronouncements are quite different to legal obligations. In their article on content governance, Bowers and Zittrain criticize the present version of internet intermediaries' self-governance as unsatisfying, "corporate customer service processes" designed to "defuse PR pressure and protect profitability," as opposed to robustly moderating harmful content.[83]

The question that then arises is to what extent assertions of responsibility should translate into legal liability, and whether we're prepared to accept poor outcomes so long as the companies demonstrate that they are making an effort.

Facebook is clearly wise to the business risk such regulation would create for them. In February 2020, Facebook proactively released a policy paper entitled *Charting a Way Forward: Online Content Regulation*, in which the tech giant's vice president of content policy acknowledged the need to discuss new regulatory frameworks to address the growing tension between the virtue of globalized free speech and the harms of globalized terrorism, trolling, and the like. Perhaps unsurprisingly, the report claims that companies like Facebook are "intermediaries not speakers," and that they should therefore be spared from liability for user-generated content they publish, promote, and recommend.[84]

According to Facebook, the appropriate response to these new online threats is for internet intermediaries to maintain appropriate and transparent "systems and procedures" (possibly with government regulation of certain benchmarks in complaint responsiveness) to balance user safety

---

82 Rocha, V. *et al*. (2018) *Mark Zuckerberg testifies before Congress.* CNN. https://www.cnn.com/politics/live-news/mark-zuckerberg-testifies-congress/h_908afd7a7eabfdc60a62e21700493e2c [Accessed May 26, 2020.]

83 *Age of Disinformation, supra*

84 Bikert, M. (2020) *Charting a way forward: Online content regulation*. Facebook. https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward_Online-Content-Regulation-White-Paper-1.pdf [Accessed May 26, 2020.] [*Charting a Way Forward.*]

with freedom of expression. At the same time, Facebook cautions that any regulation should, ". . . take into account a company's size and reach as content regulation should not serve as a barrier to entry for new competitors in the market." [85]

More recently, Facebook announced its intention to create its own "supreme court" known as the *Oversight Board*, where users can appeal the platform's decisions to remove or not remove content.[86] The move was heavily criticized as a distraction that will not be able to resolve complaints expeditiously and will serve to deflect criticism of Facebook's permissive editorial policies away from Mr. Zuckerberg.[87]

Economic Arguments Behind Limiting Liability for Internet Intermediaries

Despite evidence to the contrary presented above, Facebook's position indicates a clear preference to be treated like a billboard at law: a passive, unthinking intermediary that does nothing more than obediently transmit users' content—legal or otherwise. If we accepted their position, the *Byrne* standard of notification would apply, and this is how the platforms have crafted a series of arguments to justify their non-compliance.

Generally, platforms point to a few metrics to argue that robust content moderation is impossible:

> **1. Cost of enforcement.** At the present time, for example, Facebook's content is moderated by a workforce of 35,000, whose mandate of moderation runs the gamut from defamation to hate speech to child pornography and everything objectionable in between.[88] While the question of content moderation may be easier to determine at the extremes (e.g., child pornography or graphic depictions of violence), it is harder to discern when confronted with nuanced questions of libel or privacy rights, which often require a careful balancing of legal interests. It should be noted that while criminal offences like child pornography and violent images are often addressed by direct

85 *Charting a Way Forward, supra.*

86 At the time of writing, the *Oversight Board* has been created but is not yet operational. See https://www.oversightboard.com

87 Walther, M. (2020) Facebook's 'Supreme Court' won't solve anything. The Week. https://theweek.com/articles/913248/facebooks-supreme-court-wont-solve-anything [Accessed June 4, 2020.]

88 See, for example, Holmes, A. (2019) *The company behind Facebook's nightmarish moderation center in Florida will end its content moderation services*. Business Insider. https://www.businessinsider.com/facebook-content-moderator-cognizant-cancels-contract-2019-10 [Accessed June 4, 2020.]

intermediary-to-police communication, and highly calibrated removal technology,[89] more complex and context-specific civil claims like defamation and harassment are not dealt with as expeditiously.

2. No knowledge of the impugned content and defences to liability. Related to point one is the question of the *context* and evidentiary support for certain publications that may be complained of, but of which an intermediary may not be aware. Take for example a complaint of defamation under Canadian law, which can be protected by myriad legal defences including truth, fair comment, qualified privilege, and responsible communication. As mentioned above briefly, such defences are the true "engine" of any defamation action, and each requires an in-depth knowledge of the *background* of the publication—something of which only those parties that are intimately involved with the publication would be aware.

3. An avalanche of litigation. Facebook has billions of users but just tens of thousands of content moderators. This suggests that any definitive judicial pronouncement on their liability as publishers could spark a cottage industry in frivolous and vexatious claims against large intermediaries by plaintiffs hoping to make easy money.

In his experience, the legal consultant to this paper has found that Facebook's typical pre-litigation stance is to cite the fact that it is home to billions of pieces of content, such that policing it effectively is akin to "finding a needle in a haystack." For example, Facebook recently faced legal action in Northern Ireland for hosting a page that encouraged violence against released prisoners by listing their names and addresses and detailing their offences. In its defence, Facebook claimed, without evidence, that "with billions of posts, likes, photos, and comments added to Facebook daily, Facebook could not reasonably scour its site in hopes of finding [the] content at issue—a true needle in a haystack." The court ruled against them.[90]

There are serious problems with the "needle in a haystack" defence. As demonstrated earlier, the core value proposition of firms like Facebook and YouTube is the ability to offer users a personalized service and to offer advertisers the ability to micro-target ads based on each user's

89 Antigone, D. (2018) *New Technology to Fight Child Exploitation*. Facebook. https://about.fb.com/news/2018/10/fighting-child-exploitation/ [Accessed June 1, 2020.]

90 *G v Facebook Ireland Ltd & Anor* [2015] NIQB 11 (Feb. 20, 2015.). http://www.bailii.org/nie/cases/NIHC/QB/2015/11.html [Accessed June 1, 2020.]

preferences. In other words, finding the "needle in a haystack" is their core competency. And even if one were to accept their insistence on being treated as a billboard, should the platforms not be subject to the same principled policy imperatives and notions of responsibility that the *Byrne* decision creates for physical billboards?[91] Further, the fact that many mainstream online intermediaries already police and moderate content at significant cost suggests that the "needle in a haystack" is really about the limits of their willingness to open their pocketbooks.

At least in respect of large internet service providers and social media platforms, one observes private enterprises with strong profit motives and robust bottom lines. Yet Facebook's human moderators are notoriously poorly paid, and recent reports revealed that a considerable number suffer from PTSD after having viewed hundreds, sometimes thousands of posts depicting suicides, acts of child sexual abuse, terrorist violence, and worse.[92]

In the case of Myanmar it was revealed that Facebook employed just one Burmese-speaking content moderator based in Dublin to vet the content of four million Burmese-speaking users, despite repeated warnings to senior executives that the platform was being used to incite violence, action to suppress the hatred was not taken.[93] As reporter's assessment of the circumstances highlights the following:

> ". . . a barebones staff without the capacity to handle hate speech or understand Myanmar's cultural nuances, an over-reliance on a small collection of local civil society groups to alert the company to possibly dangerous posts spreading on the platform. All of these reflect a decidedly ad-hoc approach for a multi-billion-dollar tech giant that controls so much of popular discourse in the country and across the world."[94]

Indeed, even the United Nations was critical of Facebook for being a platform for hate speech and disinformation in Myanmar.[95] The resulting violence killed tens of thousands of people and forced nearly a million more to flee.

---

91 LCO, *supra* at page 31.
92 Newton, C. (2019) *The trauma floor: The secret lives of Facebook moderators in America.* The Verge. https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona [Accessed May 26, 2020.]
93 *McLaughlin, supra.*
94 *McLaughlin, supra.*
95 *McLaughlin, supra.*

It is clear from this analysis that many internet intermediaries have already built calibrated content moderation machines, and there is little doubt that they operate in a privileged place both socially and economically. It remains debatable whether their legal obligation matches that station.

Towards a New Duty of Care: Reconceptualizing Rights and Obligations for Internet Intermediaries

Making Online Justice Accessible

While the common law of defamation law provides helpful guidance as to whether platforms are responsible for harmful and illegal content that their services transmit, promote, and recommend, it has major limitations for the issue at hand because it requires individuals to pursue their case one at a time; each on their own individual facts. Most people lack the time, money, and determination to engage in trench warfare with the likes of Facebook and Google, and those that do encounter a war of attrition strategy that seeks to exhaust complainants' financial and emotional resources.

Besides, even if the plaintiff succeeds, the process is lengthy and the damage has usually already been done. Canadian defamation law has traditionally acknowledged that in publication, "the truth rarely catches up with the lie,"[96] and this lament is only intensified when applied to the online world. Similarly, modern academia and case law have highlighted the notion of reputation as inherently tied to personal dignity, social connection, and in turn the good of society itself.[97]

It may therefore be useful to look at the problem through the prism of social good and public health. Perhaps internet intermediaries should assume obligations to the societies around them.

A new-found legal emphasis on personal well-being in the face of internet intermediaries' power would not be alien given the present trajectory of Canadian law. The fundamental principles of "reputation" and "privacy"—principles being destabilized by unconstrained internet intermediaries—have ever-increasing primacy in a world defined by the internet.

What is more, the Supreme Court of Canada has held that privacy—and *particularly,* a child's right to privacy—goes to the heart of our fundamental liberty and security guaranteed by the *Charter*.[98] The court has also observed that privacy is no less an important *social* good, ruling that, "privacy is essential for the well-being of the individual…The restraints imposed on government to pry into the lives of the citizen go to the essence of a democratic state."[99] However, as Zuboff highlights, the

96 *Barrick Gold Corp. v. Lopehandia* (2004), 71 O.R. (3d) 416 (C.A.), at para. 32 [*Lopehandia*].
97 *LCO, supra* at page 31-32.
98 *A.B. v. Bragg Communications,* 2012 SCC 36, at paras. 14-29.
99 *R v Dyment*, [1988] 2 SCR 417, para 17 [*Dyment*] as cited by *LCO*, supra at page 31-32.

authors of our constitutions did not envision the possibility that private companies would one day pose a far greater threat to individual privacy than governments.[100]

Policymakers should therefore consider whether continuing the current hands-off approach to internet intermediaries is justifiable, given that for many, intermediaries are the most visible and arguably the most responsible for the speed and breadth of which harmful content is disseminated. Modern Canadian case law is replete with examples of litigants unable to secure meaningful justice because the original publishers are either anonymous or outside the jurisdiction. In such cases, direct resort to internet intermediaries may be the *only* reasonable avenue for a plaintiff to pursue under the circumstances.

It is worth noting the judiciary and the bar's new, albeit halting, focus on providing "access to justice" for average Canadians seems particularly prescient when considering disputes between large internet platforms and their users. The Supreme Court has noted the need for a "culture shift" in litigation so that it can be financially and practically accessible to ordinary Canadians,[101] and it is submitted here that increased attention to internet intermediary liability might be a productive start in giving voice to this principle.

What is more, encouraging meaningful citizen engagement with the legal system does not necessarily expose internet intermediaries to inappropriate levels of legal risk, frivolous lawsuits, or muzzling of their free expression interests—again, Canada's legal system already has the building blocks in place to prevent such abuses. In Ontario, for example, section 137.1 of the *Courts of Justice Act*, is an anti-strategic litigation against public participation (anti-SLAPP) provision—a legal framework for defendants to bring preliminary motions that can kill legal cases designed to chill discussion on issues of public importance. While this mechanism was initially envisioned as a way for impecunious defendants to preserve their free expression rights in the face of what would otherwise be "David vs. Goliath" legal battles, there is no reason why large internet intermediaries should not be able to avail themselves of the same legislation (perhaps with a few amendments) in order to preserve their own free expression rights.

Given the rapidly proliferating scope of internet harms that have been outlined in this paper so far,

100 Tsalikis, C., Zuboff, S. (2019) Shoshana Zuboff on the Undetectable, Indecipherable World of Surveillance Capitalism. CIGI (Interview). https://www.cigionline.org/articles/shoshana-zuboff-undetectable-indecipherable-world-surveillance-capitalism [Accessed June 4, 2020.]

101 *Hryniak v. Maudlin*, 2014 SCC 7 at paras 1, 2, 24, 27.

as well as the apparent systemic objections that intermediaries themselves have to a robust form of regulation, Canadian policymakers must decide what the proper regulatory relationship is between the individual, the internet intermediary, and the state.

The International Scene

UK lawmakers have taken a position on this question, and they have landed on the side of the public interest, reporting, ". . . as the power and influence of large companies has grown, and privately run platforms have become akin to public spaces, some of these companies now acknowledge their responsibility to be guided by norms and rules developed by democratic societies."[102]

What can be done, then, to strike an appropriate balance between the individual's right to be free from unlawful conduct online, and an internet intermediaries' and society's interest in promoting free expression?

Other democratic jurisdictions are implementing robust policy responses to online harm, though it must be noted that the legal systems in these countries are not always analogous to Canada's:

- Germany adopted the Network Enforcement Act (NetzDG) in 2017, which requires online platforms with more than two million registered users in Germany to remove "manifestly unlawful" content that contravenes specific elements of the German criminal code (for example, holocaust denial or hate speech). Platforms are obliged to remove such content within 24 hours of receiving a complaint and must remove all other 'unlawful' content within seven days of notification. Non-compliance risks fines of up to €50 million, and the legislation also seeks to increase platform responsibility through imposing transparency and reporting obligations.[103]

- Australia established an eSafety Commissioner through its Enhancing Online Safety for Children Act in 2015. The eSafety commissioner is responsible for promoting online safety for all Australians. Its mandate includes a complaints service for young people as well as removing illegal online content and responding to image-based abuse.[104]

102 *Online Harms, supra,* at *Clarity for Companies*, 13

103 *Network Enforcement Act (Netzdurchsetzunggesetz, NetzDG)*. German Law Archive. https://germanlawarchive.iuscomp.org/?p=1245 [Accessed May 26, 2020.]

104 *Our legislative functions*. eSafety Commissioner, Australia. https://www.esafety.gov.au/about-us/who-we-are/our-legislative-functions [Accessed May 26, 2020.]

- In 2018, the European Commission published a proposal on preventing the dissemination of terrorist content online, as well as an Action Plan against Disinformation with help from companies including Facebook, Google, and Twitter; [105]

- The United States Congress is currently debating the *EARN IT Act*, which proposes to strip platforms of their liability provided under Section 230 of the *Communications Decency Act,* "if they cannot prove they're doing enough to combat child exploitation."[106] The bill enjoys bi-partisan support, though for very different reasons.[107]

- As cited throughout this paper, the United Kingdom has embarked upon a significant reconceptualization of the proper relationship between large internet intermediaries and the state. The *Online Harms White Paper* is focused on governmental oversight to ensure certain regulatory and qualitative benchmarks in terms of technological effort and temporal responsiveness. It recommends the creation of a "statutory duty of care," which is discussed below.[108]

According to Bowers and Zittrain, the solution may lie in developing governance structures outside the intermediaries themselves and/or reorienting themselves not merely as socially-progressive actors, but also as "content fiduciaries"— that is to say parties with a legal duty to act in their users' best interests akin to other relationships where there is an asymmetry of power, such as doctor to patient or lawyer to client.[109]

Conceptualizing the Internet Intermediaries' Duty to Users

If one were to look at the question from a basic tort law analysis, a prospective litigant would have to demonstrate that the act or omission with which the intermediary is charged was a breach of

105 *Online Harms, supra,* at 2.16.
106 Levine, A. (2020) *EARN It Act under the Senate Judiciary microscope.* Politico. https://www.politico.com/newsletters/morning-tech/2020/03/11/earn-it-act-under-the-senate-judiciary-microscope-488556 [Accessed June 1, 2020.]
107 Democrats are seeking to limit platforms' immunity for hosting hateful or illegal content, to reduce its prevalence. However, Republicans are seeking to limit platforms' immunity for *removing* such content, arguing that legitimate conservative opinions are being censored under the guise of reducing hate speech and lies.
108 *Online Harms, supra*
109 *Age of Disinformation, supra*

the intermediary's "legal duty" to the plaintiff. In turn, the plaintiff would need to show that such a duty actually exists in law—i.e., that the relationship between the parties was of sufficient *closeness* or *reliance* that it would be reasonable to expect the intermediary to owe the plaintiff some kind of legal obligation. If a duty were found to exist, then it would need to be shown that the intermediary acted unreasonably in the exercise of that duty—in legal terms that they fell below the "standard of care."[110] This emphasis on acts or omissions means that the difference between asking whether an internet intermediary is a liable "publisher," or whether they have breached a "standard of care" is not all that different.

The scope of a particular duty of care is not so much a list of specified obligations, as it is a legal principle that is clarified, specified, and consolidated with each new legal decision. As time marches on, the scope of certain relationships become well-worn and accepted, while others like those between internet intermediaries and users remain untouched.

The basic test for determining a duty of care is what a reasonable person would do to act responsibly. For example, companies that make playpens for infants would have a duty of care to ensure that they are safe by using materials that are non-toxic and of suitable quality. Likewise, drivers of automobiles have a duty of care to other drivers—for example, to be alert and sober.

This does not mean that playpen manufacturers are responsible for every conceivable malfunction of their products. The law of tort/negligence does not hold people to standards of perfection when considering legal liability. Rather, it asks if they have acted *reasonably under the circumstances* when it comes to whether the standard of care has been breached.

The intimate informational and social nature of the relationship between internet intermediaries and their users—particularly large and powerful intermediaries with significant influence—means that a duty of care could likely be established at law. Where the litigation fireworks would emerge is in answering the question of what the appropriate *standard of care* would be for internet intermediaries in dealing with their users. In other words, once you have signed up for their service, what minimum standard are intermediaries required to meet in order to keep you safe.

The UK government, following Bowers and Zittrain,[111] are calling for a new and more detailed

110 The seminal case on the issue is *McAlister (Donoghue) v. Stevenson* (1932), [1932] A.C. 562 (U.K. H.L.).
111 *Age of Disinformation, supra*

statutory duty of care, informed and bolstered by state regulation, rather than platforms' self-professed "best practices." Much like the German framework noted above, the UK's proposals focus on how companies should behave after they are notified of the existence of harmful or illegal content on their platforms while also creating enforceable standards of practice. For example, the UK proposals would see intermediaries forced to regularly disclose the prevalence of harmful content on their platform to regulators, and to prove to regulators that their algorithms are designed to identify and remove harmful content wherever possible.[112] The proposal recommends significant penalties for companies that refuse to comply, including significant fines and liability not just for the companies but also for their senior executives, individually.[113] UK lawmakers also propose to be allowed to "disrupt" the operations of non-compliant firms, up to and including blocking them.[114]

112 *Online harms, supra,* at *Executive Summary*, 23
113 *Ibid,* at *Executive Summary*, 19
114 *Ibid,* at *Executive Summary*, 40

Recommendations for Canadian Policymakers

1. Protect freedom of expression and acknowledge its balance with other rights. In the pre-internet world, Canadians enjoyed healthy protections on free expression while also ensuring that libel, defamation, hate speech, threats, and other harmful communications were sanctioned and punished. Traditional broadcast and print publishers have held themselves to a high standard of journalistic and legal integrity, without shying away from criticizing the powerful, precisely because of Canada's robust legal system. Applying the same standard to internet intermediaries is not an act of unprecedented censorship but rather a continuation of a system that has worked relatively well. These provisions exist not to limit, but rather to *protect* the freedom of expression of certain groups who are otherwise disproportionately compelled to isolate themselves from public debate for fear of harassment or abuse. Policymakers need not prescribe which content is *appropriate*. The focus, rather, should be on the extent of the platforms' liability for content that is *already deemed inappropriate*. Today, intermediaries are censoring content, either by deciding not to present it to users or by unilaterally removing it from their platforms. Their motivations, however, cannot be separated from their economic imperatives. As argued throughout this paper, this leads to an unbalanced approach to online harms. Government regulation should focus on correcting this imbalance.

2. Acknowledge that certain online harms are categorically unacceptable. Platforms benefit from a sense of predeterminism amongst, for example, victims of cyberbullying and invasions of privacy and the resignation it instills. Such predeterminism supports their business interests but it is not a natural law. Canadian courts and legislators should question a hands-off approach in favour of a well-thought-out framework that is firmly grounded in Canadian law and democratic values, even if those diverge from a platform's business interest. Policymakers should be consoled that a more robust approach to attacking online harms is not an alien or draconian principle in Canadian law. On the contrary, and as this paper has already canvassed, the common law and statute already acknowledges the need for principled interventionism (think of *Byrne* and Manitoba's *Intimate Images* Act), while at the same time creating a nascent framework for culling frivolous or problematic claims that do not give appropriate regard to free expression rights (as seen by Ontario's Anti-SLAPP provisions).

3. Appreciate the platforms' technical prowess. The "needle in a haystack" defence belies an incomplete level of technical understanding among legislators and judges. By their own admission, platforms have advanced tools and technologies for discerning the content of their users' posts and

even their state of mind. Platforms assert these capacities when appealing to advertisers, however, when dealing with regulators, they present themselves as hapless conduits struggling to police bad actors who use their services in undetectable ways. This contradictory defence should not be deemed satisfactory.

4. Take a sovereign approach. Multinational businesses are not novel. Canada has considerable experience working with multinationals in highly regulated sectors like telecommunications, finance, resource development, aerospace, and defence. In those cases, policymakers have insisted that companies follow Canadian laws and rules if they are to continue operating in Canada. Governance of companies operating in the online sphere should be approached similarly, regardless of how normalized certain behaviours have become as a result of the laws of other jurisdictions, such as Section 230 of the *Communications Decency Act* in the United States. Canada is a sovereign country, and its public policies can and should be Canadian.

5. Use enshrined legal principles as a starting point for appropriate regulation. *Byrne* sets a clear standard for liability which the internet intermediaries appear to meet. If harmful or illegal content is found on their platforms, they should be held appropriately responsible for it. Any legal/policy prescription of responsibility—whether statutory or otherwise—is not an alien concept that necessarily harms Canadian society's interest in robust free expression. Rather, it would simply be the next logical frontier of regulation that takes into account certain empirical realities that these services:

a. Have the ability to comprehend the substance of the content a user posts before it is posted; and

b. Do not simply publish but also promote and recommend content.

6. Ensure the onus does not fall on individuals. While the law of defamation provides helpful guidance as to when an intermediary should be considered complicit in the publication of harmful material, it is unreasonable to expect each and every harmed individual to initiate and finance a protracted lawsuit.[115] Even if they could, the result would be an avalanche of litigation that would clog up the courts. A different, more equitable and efficient system may be required, such as a

115 Roger McNamee, an early Facebook investor turned anti-platform advocate, argues for "new legislation to give users the right to sue for damages if they have been harmed as a result of using an internet platform." McNamee, R. (2020) *Facebook Cannot Fix Itself. But Trump's Effort to Reform Section 230 Is Wrong*. Time. https://time.com/5847963/trump-section-230-executive-order/ [Accessed June 4, 2020.]

specialized tribunal for mediating disputes between users and internet platforms. Crucially, this system must harness state power to level the playing field between complainants and the platforms. Consider a mechanism for people to file complaints with a government agency that could take investigative and enforcement actions on citizens' behalf, but with the full weight of the state and the law behind them. The Privacy Commissioner's office works in this manner, as does Manitoba's process for addressing the dissemination of intimate images.

7. Apply meaningful and proportionate sanctions. As demonstrated, intermediaries are most likely to become concerned about harmful content when they face a clear and imminent reputational or economic penalty. Any policy must provide for outcomes that effectively deter inaction and proportionally fit the magnitude of the harm that is being addressed. Without these provisions, intermediaries will likely consider the fines to be a cost of doing business, and non-compliance will persist.

8. Create and enforce strict privacy protections that minimize intermediaries' ability to collect personal data. For platforms, the "engagement imperative" is driven by two complementary forces: the drive to collect as much observational data as possible and the drive to use that data to show users as many advertisements as possible. This imperative is what incentivizes the spread of extreme, incendiary content. If this hostile business model is reined in, the resulting harms could be greatly reduced.

9. Move promptly. Many countries make public policy by copying others. If Canada is to enjoy the protection of a legal regime that reflects its needs and values, it must move quickly. Canada is uniquely placed, given its technological clout and robust rule of law, to be a world leader in addressing the very important challenges outlined in this paper. If Canada lags behind, a common consensus may emerge that does not conform to Canadians' priorities or values. It is better to help shape such a standard than to be subject to it.

The internet may be a new and rapidly evolving medium for publication, however, it is *still* a medium for publication. It is up to Canadian policymakers to both appreciate its myriad benefits while at the same time meeting its challenges and unintended social harms head-on. Creating an appropriately balanced regulatory regime for internet platforms will not be easy, nor will it be perfect. However, Canada was founded on the principle of equitable rule of law, and it is not acceptable to shy away from difficult policy debates simply because they are difficult. Writing in the

internet's infancy in 2004, the Ontario Court of Appeal cautioned against seeing the internet's size and speed as a challenge which made proper regulation impossible [citations omitted]:

> "The highly transmissible nature of the tortious misconduct at issue here is a factor to be addressed in considering whether a permanent injunction should be granted. The courts are faced with a dilemma. On the one hand, they can throw up their collective hands in despair, taking the view that enforcement against such ephemeral transmissions around the world is ineffective, and concluding therefore that only the jurisdiction where the originator of the communication may happen to be found can enjoin the offending conduct. On the other hand, they can at least protect against the impugned conduct re-occurring in their own jurisdiction. In this respect, I agree with the following observation. . .
>
> Any suggestion that there can be no effective remedy for the tort of defamation (or other civil wrongs) committed by the use of the Internet (or that such wrongs must simply be tolerated as the price to be paid for the advantages of the medium) is self-evidently unacceptable.[116]

Some sixteen years later, this dictum is even more noteworthy. Yet, it is not enough to simply profess that internet intermediaries are not above the law, for they are a law unto themselves until such time as sovereign nations like Canada assert otherwise.

116 *Lopehandia, supra,* at para. 75.