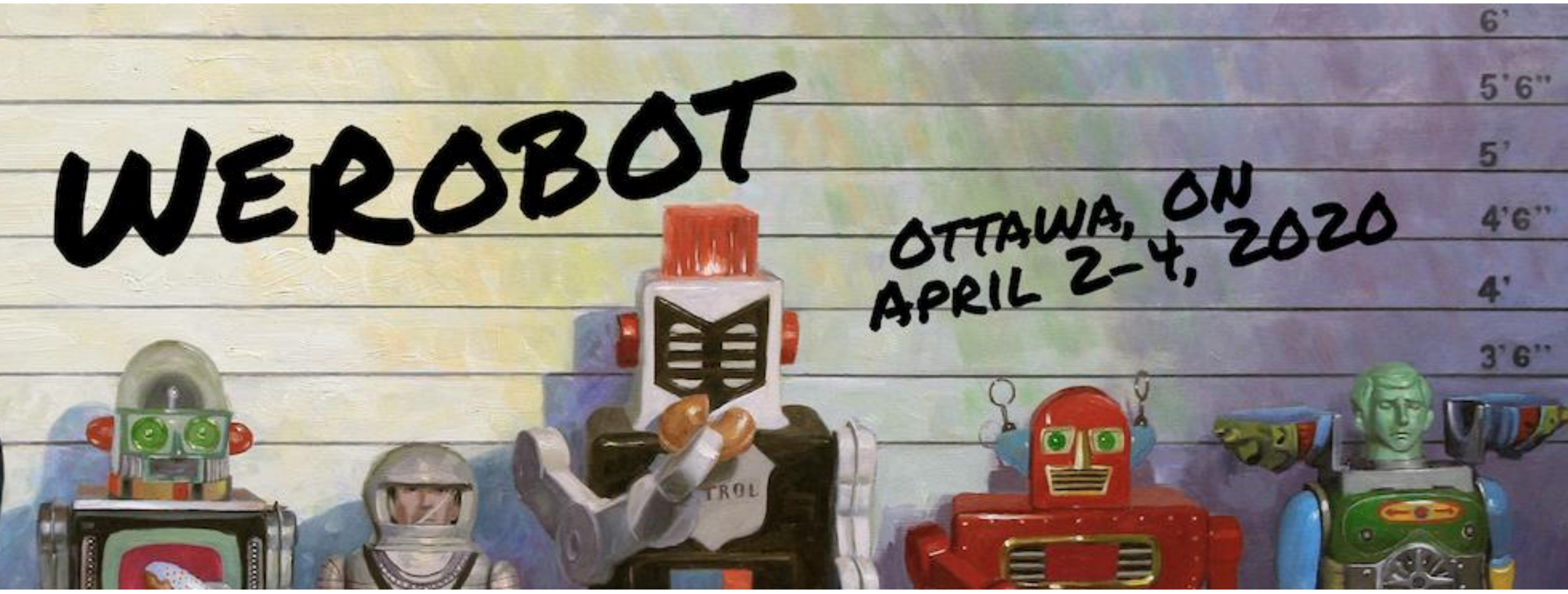

Missing the Unintended Forest despite the Deliberately Planted Trees: Reasonable Foreseeability and Legal Recognition of Platform Algorithm- Facilitated Emergent Systemic Harm to Marginalized Communities

Cynthia Khoo

University of Ottawa

Draft prepared for discussion at We Robot 2020.

Not for attribution or further circulation without permission of the author(s).

Missing the Unintended Forest despite the Deliberately Planted Trees: Reasonable Foreseeability and Legal Recognition of Platform Algorithm-Facilitated Emergent Systemic Harm to Marginalized Communities

Cynthia Khoo[†]

Draft Paper Presented at We Robot 2020 (April 3-4, 2020, Ottawa, ON)[‡]

Digital platforms such as Facebook, Google, Twitter, YouTube, Airbnb, and Uber algorithmically govern increasingly large swaths of our private lives, and have accrued discomfiting levels of power over much of public life. They have fundamentally impacted critical issues such as electoral integrity, affordable housing and public transit, and the nature of public discourse, in a way that disproportionately harms historically marginalized groups. Such harms may amount to violations of the right to equality and freedom from discrimination, of groups protected under human rights laws. While much legal and related interdisciplinary scholarship regarding platform regulation has focused on regulatory responses, the field has also left open the way for a private right of action, required as a backstop where regulation does not suffice, and to provide remedy to victims where harm has already occurred.

This paper proposes several preliminary building blocks towards developing that private right of action, beginning with establishing as a legal harm, based on human rights and environmental law principles, platform algorithm-facilitated emergent systemic harm to marginalized groups. The core concept proposed is that of emergent systemic harm, drawing on the notion of emergence from artificial intelligence and robotics literature: systemic harm which emerged unexpectedly from a complex system, whose constituent components did not in isolation necessarily raise legal or ethical concerns, nor individually indicated foreseeable or predictable harm, but which worked together in such a way as to produce a harmful result that was more than the sum of the system's parts. This paper identifies five key contributing components to emergent systemic harm, including: the platform's algorithms, the platform's design and technological affordances, the platform's business model, the time and scale of the platform's reach, and users' human nature such as individualistic and instrumental use of the platform.

Analyzing known systemic harms from online platforms that have been deemed "unintended consequences" reveals distinctions between harms that were "unintended", "unforeseen", and "unforeseeable", with further differences between what platform companies could foresee and what outside experts, internal employees, and vulnerable users foresaw—this discrepancy is termed the platform foreseeability gap. This paper then advocates for applying tort liability through negligence to platform-facilitated emergent systemic harm to marginalized groups, using an interpretation of the reasonable foreseeability element that encompasses both intersectional foreseeability and relational foreseeability, based on a contemporary understanding of the reasonable person. The final section examines how the foreseeability analysis changes in light of unsupervised autonomous algorithms and truly emergent systemic harms, and proposes a sliding scale of liability, including a no-fault insurance pool, a modified negligence test, and strict liability, depending on initial level of algorithmic risk and the actual harm that occurred.

[†] Cynthia Khoo is a technology and human rights lawyer and researcher based in Ottawa, ON. She founded a sole practice digital rights law firm, Tekhnos Law, and is a research fellow at the Citizen Lab (Munk School of Global Affairs & Public Policy, University of Toronto). Cynthia holds an LL.M. (Concentration in Law and Technology) from the University of Ottawa, where she interned as a research student and junior counsel at the Samuelson-Glushko Canadian Internet Policy and Public Interest Clinic (CIPPIC). She serves on the board of directors of Open Privacy and holds a J.D. from the University of Victoria, Faculty of Law, in British Columbia, Canada.

[‡] With immense gratitude to my LL.M. research co-supervisors: Professor Jane Bailey at the University of Ottawa Faculty of Law and co-principal investigator of The eQuality Project, and the late Professor Ian Kerr, Canada Research Chair in Ethics, Law and Technology at the University of Ottawa Faculty of Law (cross-appointed with the faculties of Medicine, Philosophy, and Information Studies). This current draft benefited significantly from Professor Bailey's valuable and much appreciated commentary on an earlier version written during the course of my LL.M. program. I am also grateful to Professor Kerr, who received my initial proposal of "emergent systemic harm" with the resounding enthusiasm for which he was so well known, and it remains a great sadness that he was unable to witness the eventual fruition of this work that he so inspiringly encouraged.

Table of Contents

Introduction

Cassandra: "And it's all the same if nothing of mine persuades you, of course: the future will come; and you will soon be at my side to pity and call me too true a prophet."
[...]
Chorus: "Still your tongue, you wretched woman! Say nothing inauspicious!"[1]

~

"I don't say 'listen to trans women' because we've got some magical 6th sense, we can't see the future. It's far, far simpler than that. We know because we are usually the first they target…"[2]

A. Overview

If a systemic harm to a marginalized group flows from an online platform, but nobody intended it, does the harm legally exist? This paper answers that question in the affirmative, through the concept of platform algorithm-facilitated emergent systemic harms to marginalized groups.

An emergent systemic harm is a type of harm that arises from the constituent parts of a complex system working together in a way that produces results—in this case, systemic harm—that are greater than the sum of its parts. Specifically, emergent systemic harm occurs when each constituent part of the system was not in itself inherently or necessarily wrong, unethical, illegal, or even unreasonable, but the parts working in conjunction with each other somehow give rise to a systemic harm usually only realized after the fact. This paper will demonstrate that popular online platforms, such as Facebook, Google, Twitter, Airbnb, Uber, or YouTube, contain several key components contributing to emergent systemic harms, including emergent systemic harms to marginalized communities (both users and non-users). These components were drawn from analyzing known harmful yet "unintended" consequences that have flowed from these platforms and users' activities on them in recent years, and are meant to provide a guide to identifying potential future emergent systemic harms from these same platforms in the future, as well as from other or new platforms that share similar constituent components.[3]

This paper also proposes holding online platforms liable in negligence, under Canadian tort law, for platform-facilitated emergent systemic harms to marginalized communities. The increasing body of academic literature, legal scholarship, and policy discourse calling for regulation of digital platforms has left open the way for a private right of action, to serve as a backstop where regulation fails and significant

[1] Aeschylus, *Oresteia*, Christopher Collard, trans. (New York: Oxford University Press, 2002) at 35 (lis 1239-47).

[2] @CaseyExplosion, "I don't say 'listen to trans women' because we've got some magical 6th sense, we can't see the future. It's far, far simpler than that. We know because we are usually the first they target, and transphobia is a bigotry that is not only acceptable, but applauded, so nobody cares." (23 May 2018 at 2:13), online: *Twitter* <https://twitter.com/CaseyExplosion/status/999352635810557954>.

[3] An earlier version of this paper focused on the components of platform-facilitated emergent systemic harms on an equal basis, where algorithms were one of several components examined alongside each other. The term platform-algorithm-facilitated emergent systemic harm simply puts the focus on algorithms front and centre and confines the focus of discussion to platform systems that involve algorithms in a contributing role, while implying that the harm was primarily due to the algorithm, as opposed to the algorithm interacting with other, equally "culpable" components of a platform system. It is theoretically possible to have a platform-facilitated emergent systemic harm arising from an online platform that does not use algorithms, but gave rise to systemic harm nonetheless based on interactions between its non-algorithmic components. However, the nature of algorithms, particularly the more autonomous and unsupervised they are, likely increases the possibility of emergent systemic harm. Throughout this paper, the term "platform-facilitated" emergent systemic harm may be used interchangeably with "platform algorithm-facilitated" emergent systemic harm, with the understanding that the emergent systemic harms discussed in this paper all involved algorithms as a key contributing factor in some way.

harm to vulnerable individuals and communities nonetheless occurs. Such an action would be distinguished from those barred by section 230 of the United States *Communications Decency Act*[4] ("CDA 230") because it focuses not on holding a platform accountable for the actions of any of its users or for any individual harms, but for the platform company's own direct actions in negligently developing, fostering, or upholding the impugned platform system itself giving rise to the identified *systemic* harm. The paper identifies negligence as an appropriate cause of action due to the rhetoric of "unintended consequences" strongly militating towards the legal framework used to address unintentional wrongdoing and unintended injury to innocent parties, in addition to other relevant benefits of applying tort law.

Closer examination of some platform-facilitated emergent systemic harms demonstrates that the "emergent" nature of such harms can become tenuous and unstable. Especially in the context of harm to marginalized communities, harm that was "emergent" or "unexpected" for some individuals may in fact have been "foreseeable" or even "predictable" for others. This disparity leads to what may be considered the "platform foreseeability gap". This emergent, quasi-emergent, or pseudo-emergent nature of some platform-facilitated systemic harms has implications for discussing and applying the reasonable foreseeability element of negligence, where reasonable foreseeability is a central component of determining duty of care, standard of care, and the defence of remoteness. In response, this paper argues that reasonable foreseeability, in Canadian law, inherently implies or ought to imply *intersectional foreseeability* and *relational foreseeability*, in part based on a critical intersectional formulation of the reasonable person.

The paper will conclude by exploring potential approaches to determining liability for platform-facilitated systemic harms that involve unsupervised autonomous algorithms, leading to harms that are more likely to be truly emergent and genuinely unforeseeable by anyone. A sliding scale approach based on level of apparent risk and actual harm is experimentally proposed as a starting point for discussion. This scale begins with a no-fault insurance pool for low- and medium-risk algorithms that resulted in low-level harm, through to a modified negligence standard for low and medium-risk algorithms resulting in medium- or high-stakes harm, and ending with strict liability for high-risk algorithms, regardless of the level of actual harm.

B. Situating Paper in Context

The topic of platform regulation has garnered increasing attention in recent years from scholars, researchers, regulators, legislators, politicians, and the public alike, with an ever-growing body of relevant academic literature, legal scholarship, policy research, and public writing.[5] Further leading work spans

[4] *Communications Decency Act*, 47 USC § 230(c)(1) (1996).

[5] See e.g., Luca Belli & Nicolo Zingales, eds, *Platform regulations: how platforms are regulated and how they regulate us. Official outcome of the UN IGF Dynamic Coalition on Platform Responsibility*, 1st ed (Rio de Janeiro: Escola de Direito do Rio de Janeiro da Fundação Getulio Vargas, 2017); Tarleton Gillespie, "Platforms Are Not Intermediaries" (2018) 2 Geo L Tech Rev 192 [Gillespie, "Platforms"] (articulating the collective public impact of users' individual actions on platforms); Kate Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech" (2018) 131 Harvard L Rev 1598 (documenting and analyzing online platforms as "the new governors", including parallels between U.S.-based social media companies' internal content moderation policies and U.S. First Amendment jurisprudence); Danielle Citron & Quinta Jurecic, "Platform Justice: Content Moderation at an Inflection Point" (2018), online (pdf): *Hoover Institution* <www.hoover.org/sites/default/files/research/docs/citron-jurecic_webreadypdf.pdf>; Frank Pasquale, "Internet Nondiscrimination Principles: Commercial Ethics for Carriers and Search Engines" (2008), online (pdf): *University of Chicago Legal Forum* <chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1427 &context=uclf> [Pasquale, "Search Engines"]; Frank Pasquale, "Two Narratives of Platform Capitalism" (2016) 35 Yale L & Pol'y Rev 309; Frank Pasquale, "Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic

across additional and adjacent fields and disciplines, including: media and communications studies, science and technology studies, (techno-)sociology,[6] critical race and technology studies,[7] communications law and policy,[8] intermediary liability law and policy,[9] human rights law (in particular, concerning the rights to equality, freedom of expression, and privacy), competition and consumer protection law, and content moderation.[10]

This paper has also been greatly influenced and informed by legal scholarship that does not fall directly into the field of platform regulation *per se*, but which formed the foundation of much of the analytical work done in this paper to extend pre-existing legal doctrines and legal theory to the context of digital platforms. This included, for example, Jane Bailey's work on technology-facilitated gender-based violence, as an example of how technology can impact the right to equality;[11] Ian Kerr's conceptualization of relational privacy,[12] based on Carys Craig's conceptualization of relational copyright;[13] Marina Pavlovíc's scholarship on platform-to-consumer contracts (also known as terms of

Society" (2017) 78:5 Ohio St LJ 1243; and Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019) (exhaustively detailing how platforms' data collection strategies and associated digital infrastructure have given rise to a new socioeconomic technopolitical world order that threatens human autonomy and dignity on an existential level); and Mike Masnick, ed, *Techdirt*, online: <www.techdirt.com/>.

[6] See e.g., Zeynep Tufekci's public and scholarly writing documenting the radicalizing effects of ad-driven algorithms: Zeynep Tufekci, "YouTube, the Great Radicalizer", *The New York Times* (10 March 2018), online: <www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html> [Tufekci, "Great Radicalizer"]; and Zeynep Tufekci, "Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency" (2015) 13 Colo Tech LJ 203 [Tufekci, "Computational Agency"].

[7] See e.g., Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018) (detailing the impacts of Google's search algorithms on the public information environment and the harmful consequences to members of racialized communities, particularly black women and girls).

[8] See e.g. Harold Feld, "The Case for the Digital Platform Act: Market Structure and Regulation of Digital Platforms" (May 2019), online (pdf): *Public Knowledge* <www.publicknowledge.org/download/the-case-for-the-digital-platform-act/> (applying lessons from traditional and contemporary communications law to proposing a holistic legal framework for platform regulation).

[9] See e.g., the work of Daphne Keller ensuring clarity around the invocation and application of intermediary liability law: Daphne Keller, "Who Do You Sue? State and Platform Hybrid Power over Online Speech" (29 January 2019), online (pdf): *Hoover Institution* <www.hoover.org/sites/default/files/research/docs/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech_0.pdf>. See also generally "Intermediary Liability", online: *Center for Internet and Society* <cyberlaw.stanford.edu/focus-areas/intermediary-liability>.

[10] See e.g., Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (New Haven, CT: Yale University Press, 2018) [Gillespie, *Custodians*] (arguing content moderation is not a secondary task or necessary afterthought but in fact the core service that online platforms provide to users); Sarah T Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media* (New Haven: Yale University Press, 2019); and Robyn Caplan, "Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches" (14 November 2018), online (pdf*): Data & Society* <datasociety.net/wp-content/uploads/2018/11 /DS_Content_or_Context_Moderation.pdf> (delineating how differently scaled and oriented digital platforms engage in different approaches to content moderation).

[11] See e.g. Jane Bailey & Carissima Mathen, "Technologically-Facilitated Violence Against Women & Girls: Assessing the Canadian Criminal Law Response" (2020) Can Bar Rev 564.

[12] Ian Kerr, "Schrödinger's Robot: Privacy in Uncertain States" (2019) 20:1 Theor Inq L 123 [Kerr, "Schrödinger's Robot"]

[13] Carys Craig, *Copyright, Communication and Culture: Towards a Relational Theory of Copyright Law* (Cheltenham: Edward Elgar Publishing, 2011) [Craig, *Relational Theory*]

service)[14] and their heightened stakes for a citizenry forced to be "consumers first";[15] and Leslie Bender's, Lynda Collins', and Heather McLeod-Kilmurray's scholarship regarding environmental tort law and (eco)feminist tort theory.[16] Furthermore, given the focus on platform algorithms in particular, this paper also relies on scholarship in artificial intelligence (AI) and robotics law and policy, including being guided by the values and understanding encompassed by work advancing algorithmic fairness, accountability, transparency, and ethics (FAT(E)),[17] drawing on Ryan Calo's articulation of emergence in the context of algorithms,[18] and looking to legal scholarship analyzing potential ways to apply legal liability to artificial intelligence and machine learning algorithms in the event of accidental injuries.[19]

As mentioned, much of the literature, research, and scholarship regarding platform regulation to date has concerned what the term suggests: regulation. Recommendations for imposing constraints and accountability on digital platforms[20] and their excesses of power often look to regulatory measures and frameworks meant to prevent or mitigate harm pre-emptively, or look to indirect solutions such as increasing competition, rather than apply direct legal liability as in the case of other entities that cause legally actionable harm. There have been and continue to be many good reasons for this approach, nearly all of which are encompassed in the principles of intermediary liability, an area of technology law and policy centrally concerned with not holding intermediary entities liable for the wrongdoing of end-users, for fear of quashing innovation, technological progress, and the many sociopolitical benefits that have

[14] Marina Pavlovíc, "Contracting Out of Access to Justice: Enforcement of Forum Selection Clauses in Consumer Contracts" (2016) 62:2 McGill LJ 389 [Pavlovíc, "Contracting Out"].
[15] Marina Pavlovíc, "Consumer rights in a radically different marketplace" (4 June 2018), online: *Policy Options* <policyoptions.irpp.org/magazines/june-2018/consumer-rights-radically-different-marketplace/> [Pavlovíc, "Consumer rights"].
[16] See e.g. Heather McLeod-Kilmurray, "An Ecofeminist Legal Critique of Canadian Environmental Law: The Case Study of Genetically Modified Foods" (2009) 26 Windsor Rev Legal Soc Issues 129 [McLeod-Kilmurray, "Ecofeminist Legal Critique"]; Lynda M Collins, "Material Contribution to Risk in the Canadian Law of Toxic Torts" (2016) 91:2 Chicago-Kent L Rev 567; Meinhard Doelle, "The Canadian Law of Toxic Torts, by Lynda Collins & Heather Mcleod-Kilmurray" (2015) 52 Osgoode Hall LJ 1151; and Leslie Bender, "A Lawyer's Primer on Feminist Theory and Tort" (1988) 38:1/2 J Leg Educ 3.
[17] See e.g. Brent Daniel Mittelstadt et al, "The ethics of algorithms: Mapping the debate" (2016) Big Data & Soc 1; "Critical Algorithm Studies: a Reading List" (last updated 15 December 2016), online: Social Media Collective <socialmediacollective.org/reading-lists/critical-algorithm-studies/>; AI Now Institute, "AI Now Law and Policy Reading List" (1 October 2018), online: *Medium* <medium.com/@AINowInstitute/ai-now-law-and-policy-reading-list-641368f09228>.
[18] See e.g. Ryan Calo, "Robotics and the Lessons of Cyberlaw" (2015) 103:3 Cal L Rev 513 [Calo, "Lessons"]; Ryan Calo, "Robots in American Law" (2016) online (pdf): <euro.ecom.cmu.edu/program/law/08-732/AI/Calo.pdf> [Calo, "Robots"].
[19] See e.g. Yavar Bathaee, "The Artificial Intelligence Black Box and the Failure of Intent and Causation" (2018) 31:2 Harvard JL & Tech 890; Hannah R. Sullivan and Scott J. Schweikart, "Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused by AI?" (2019) 31:2 AMA J Ethics 160; Brandon W Jackson, "Artificial Intelligence and the Fog Of Innovation: A Deep-Dive on Governance and the Liability of Autonomous Systems" (2019) 35:4 Santa Clara High Tech LJ 35; Paulius Cerka, Jurgita Grigene & Gintare Sirbikyte, "Liability for damages caused by artificial intelligence" (2015) Computer L & Sec Rev 1.
[20] "Digital platforms" and "online platforms" are used interchangeably throughout this paper and are identified using the general definition set out by Gillespie: "online services that a) host, organize, and circulate users' shared content or social interactions for them, b) without having produced or commissioned (the bulk of) that content, c) built on an infrastructure, beneath that circulation of information, for processing data for customer service, advertising, and profit." Gillespie, *Custodians*, *supra* note 10 at 18. This includes social media platforms such as Facebook, Instagram, Reddit, and Twitter; search platforms such as Google; "gig economy" service platform such as Airbnb, Uber, and Lyft; review and advice platforms such as Yelp and Trip Advisor; and e-commerce platforms such as Amazon and eBay. While this paper also follows Gillespie in excluding private messaging apps such as WhatsApp and Telegram, such apps would still be worth analyzing under the framework of emergent systemic harm proposed in this paper, where they share relevant constituent components with the digital platforms examined here.

resulted from online platforms since their introduction,[21] including for marginalized and equality-seeking groups and social movements.[22]

New developments in recent years have caused many to begin questioning, or to question with increasing force, the wisdom of continuing to shield online platforms from liability for a wide range of significant harms which platform companies may not have directly committed, but which they facilitated and profited from. The purpose of this paper is to provide some preliminary building blocks towards potential future applications of legal liability to online platforms in certain cases, in the event the courts or legislators begin moving in that direction. Namely, this paper will establish and define the concept of emergent systemic harm, and then establish the more specific concept of platform-facilitated emergent systemic harm to historically marginalized communities as a ground for legal action in Canadian law. That discussion will be followed by a multipart analysis of the notion of "foreseeability", both in the plain-language sense of how "foreseeable" certain platform-related unintended consequences may or may not have been, and in the legal sense of applying the reasonable foreseeability element of the tort law test for negligence, to platform-facilitated emergent systemic harm to marginalized groups. This section of the paper will establish two further building blocks, including the concept of the "platform foreseeability gap" and a tentative sliding scale framework to determine what standard of tort liability to apply where a platform algorithm was both unsupervised and autonomous, increasing the likelihood that a harm that occurred was genuinely unforeseeable.

The rationale behind developing platform-facilitated emergent systemic harm to marginalized groups as a legally recognized harm in Canada is as follows.[23] First, the relevant literature on platform regulation has acknowledged that there is a needed role for a private right of action as a backstop to regulation, and to provide redress to victims where harm has already occurred.[24] The Canadian legal context is an ideal one in which to shape such potentially novel actions, due to the lack of an operational Canadian equivalent of

[21] See e.g. "Manila Principles on Intermediary Liability", online: <www.manilaprinciples.org>.

[22] In this paper, the terms "marginalized", "vulnerable", or "equality-seeking" individual, group, or community, along with "historically oppressed", "historically marginalized", or "systemically oppressed" individual, group, or community will be used interchangeably. Such groups may be identified in part through their designation as protected groups under Canadian equality and non-discrimination law, in the form of federal and provincial human rights legislation and in common law jurisprudence under section 15 of the *Canadian Charter of Rights and Freedoms*, as well as in court recognition of marginalized and vulnerable groups who are subjected to systemic discrimination in other contexts, such as under search and seizure laws in a criminal justice context. Where Canadian law has not already explicitly recognized a particular class of people as having been historically and systemically oppressed so as to warrant protection under equality and non-discrimination laws, social sciences evidence may be used to make such a case in context of advocacy or litigation.

[23] It is acknowledged that most of the platform companies discussed in this paper are based in the United States, as is much of the platform-related scholarship cited. However, this paper remains relevant to and applicable in the Canadian context in several ways, beyond the fact that the legal analysis is rooted nearly entirely in Canadian law. First, arguments presented in the context of platform regulation in the United States may be transferrable to the Canadian context or otherwise highlight gaps where Canadian law is lacking (or vice versa, where Canadian law is already an improvement). Second, where there is no equivalent law, policy, or research in Canada on a specific point or issue, experiences, evidence, and systems from other jurisdictions still offer a model and jumping off point from which to begin evaluating how Canadian law ought to approach platform regulation and platform-facilitated emergent systemic harms. Third, while many of the largest platform companies are based in the United States, Canadian law nonetheless governs entities that have a "real and substantial connection" to the country (*Club Resorts Ltd v Van Breda*, 2012 SCC 17) which it may be easily argued these companies do (such as by reference to decisions involving Google and Facebook, issued by the Office of the Privacy Commissioner of Canada), even where they do not already have Canadian offices or subsidiaries. Fourth, future digital platforms that facilitate emergent systemic harms may be Canadian businesses based in Canada, and thus the legal arguments this paper provides, based in Canadian law, would directly apply.

[24] See e.g. Feld, *supra* note 8 at 118.

CDA 230;[25] a more balanced view of the right to freedom of expression, where section 2(b) of the *Canadian Charter of Rights and Freedoms*[26] is one of many human rights to be balanced against each other, as opposed to the near-absolute nature of the First Amendment in the United States; and judicial precedents grappling with the notion of intermediary liability[27] while also holding online platforms accountable in various contexts, such as in a privacy class action,[28] an intellectual property injunction,[29] and various defamation cases.[30] If there is to be a private right of action against online platforms, plaintiffs, their counsel, and the courts must be able to identify what the harm grounding the action is, as well as the causal elements of that harm. In addition, marginalized communities specifically require protection in the context of digital platforms, in the exact same way the law grants them particular protection under section 15 of the *Charter* and statutory equality and non-discrimination laws, and given the disproportionate impact that platform-related harms have had on equality-seeking groups.[31] The hope is that delineating the existence and characteristics of platform-facilitated emergent systemic harm to marginalized communities will encourage earlier recognition of such harms in future cases, before they fully come to fruition, and prompt mitigatory, preventative, or remedying responses accordingly, rather than merely identifying harms and their causes after the fact, when the damage to already vulnerable and historically oppressed communities has already been done.

C. Structure of the Paper

Part I of this paper will introduce and define the concept of emergent systemic harm, drawing on Canadian human rights and equality jurisprudence to define systemic harm, and using relevant literature

[25] While the *Canada-United States-Mexico Agreement* (CUSMA, also known as USMCA or T-MEC) imported a version of CDA 230 into Canadian law, the agreement has yet to be ratified and implemented in domestic law. *Canada-United States-Mexico Agreement*, 30 November 2018, at Article 19.17 (not yet ratified) [*CUSMA*].

[26] *Canadian Charter of Rights and Freedoms*, s 15(1), Part I of the *Constitution Act, 1982*, being Schedule B to the *Canada Act 1982* (UK), 1982, c 11 [*Charter*].

[27] *Crookes v Newton*, 2011 SCC 47.

[28] *Douez v Facebook, Inc*, 2017 SCC 33; see also various decisions from the Office of the Privacy Commissioner of Canada regarding online platforms and their privacy obligations to users: *Use of sensitive health information for targeting of Google ads raises privacy concerns* (14 January 2014), PIPEDA Report of Findings #2014-001, online: Office of the Privacy Commissioner of Canada <www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2014/pipeda-2014-001>; *Google Inc. WiFi Data Collection* (6 June 2011), PIPEDA Report of Findings #2011-001, online: Office of the Privacy Commissioner of Canada <www.priv. gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2011/pipeda-2011-001/>; *Company's re-use of millions of Canadian Facebook user profiles violated privacy law* (12 June 2018), PIPEDA Report of Findings #2018-002, online: Office of the Privacy Commissioner of Canada <www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2018/pipeda-2018-002/>; *Joint investigation of Facebook, Inc. by the Privacy Commissioner of Canada and the Information and Privacy Commissioner for British Columbia* (25 April 2019), PIPEDA Report of Findings #2019-002, online: Office of the Privacy Commissioner of Canada <www.priv. gc.ca/en/opc-actions-and-decisions/investigations /investigations-into-businesses/2019/pipeda-2019-002/> [*Joint investigation of Facebook*].

[29] See e.g. *Google Inc v Equustek Solutions Inc*, 2017 SCC 34.

[30] See e.g. Sue Gratton, "Defamation Law in the Age of the Internet: Consultation Paper" (November 2017), online (pdf): *Law Commission of Ontario* <www.lco-cdo.org/wp-content/uploads/2017/12/Defamation-Consultation-Paper-Eng.pdf>, at 102.

[31] See e.g. See e.g. Dottie Lux & Lil Miss Hot Mess, "Facebook's Hate Speech Policies Censor Marginalized Users", *Wired* (14 August 2017), online: <www.wired.com/story/facebooks-hate-speech-policies-censor-marginalized-users/>; Gaby Hinsliff, "Airbnb and the so-called sharing economy is hollowing out our cities", *Guardian* (31 August 2018), online: <www.theguardian.com/commentisfree/2018/aug/31/airbnb-sharing-economy-cities-barcelona-inequality-locals>; and "Toxic Twitter: Violence and Abuse against Women Online" (2018), online (pdf): *Amnesty International* <www.amnestyusa.org/wp-content/uploads/2018/03/Toxic-Twitter.pdf>.

from artificial intelligence and robotics law and literature to define the concept of emergence. The defining characteristic of *emergent* systemic harm is that such harm emerges from the interactions of components in a particular system—in this case, online platforms—where each component individually would not appear to give rise to harm. Part I will discuss key constituent components likely to work together to give rise to platform algorithm-facilitated emergent systemic harm: a) the platform algorithms themselves; b) the platform's design and technological affordances; c) the platform's business model; d) time and scale (of the platform's existence, reach, and user base); and e) user behaviour, including human nature in general. This discussion will draw on real-world examples to demonstrate how each component operates individually and in a way that can contribute to emergent systemic harm.[32]

Part II will argue that emergent systemic harm to marginalized groups in particular can and should constitute a legally actionable harm in Canadian law. This is based on Canadian equality and non-discrimination law recognizing that members of protected groups may bring action if they have been subjected to adverse effects of seemingly "neutral" systems that in fact contribute to systemic discrimination. Support for recognizing platform-facilitated emergent systemic harm to marginalized groups can also be found in Canadian environmental law, which grants legal protection and remedy to system-level, collective interests. Connections between environmental law and platform-facilitated emergent systemic harm are furthered by discursive comparisons of the Internet and digital platforms to their own form of ecologies and ecosystems, as well as in ecofeminist tort theory and feminist environmental legal theory proposing a more expansive and equitable version of various elements of tort, such as reasonable foreseeability and standard of care.

Part III more closely scrutinizes specific examples of platform-facilitated emergent systemic harms alongside the associated discourse of "unintended consequences" frequently applied to such harms, in order to inform how the foreseeability analysis in the negligence test would apply in this context. This section of the paper demonstrates that in many cases, what at first appeared to be an emergent systemic harm would more appropriately be considered only a quasi-emergent or pseudo-emergent systemic harm. This is because upon closer examination, the systemic harm that occurred did not arise from a system where *no* components or actors were acting entirely legally, ethically, and reasonably. Rather, what first appeared to be an unintended—implied as unforeseen—consequence, or "emergent" systemic harm, could be traced back to some incriminating element in the timeline or system, such as documented or reasonably imputed knowledge combined with continued inaction on the part of the platform company. This observation leads into an analysis distinguishing between harms that are "unintended", versus "unforeseen" in fact (by the platforms, but foreseen by others), versus "unforesee*able*" (and by whom). The result of this discussion is crystallization of what I term the "platform foreseeability gap": the discrepancy between what vulnerable and marginalized individuals and communities (as well as their abusers) can foresee as risks and harms arising from a digital platform's system features, and what risks and harms the platform company itself is ostensibly able to foresee, if any.

Part IV provides a deep dive into the foreseeability element of the legal test for negligence in Canadian tort law, in the context of platform algorithm-facilitated emergent systemic harm to marginalized communities. This section first elaborates on the suitability of applying tort liability and negligence specifically to this class of harms. This is followed by an analysis of how various considerations under the traditional legal analysis for reasonable foreseeability might apply in the context of an online platform-facilitated emergent systemic harm. This analysis takes into account the earlier destabilization of the

[32] Examples include Google's search engine auto-suggesting and ranking high search results for harmful stereotypes against racialized persons, particularly black women and girls and YouTube's recommendation algorithms systematically leading viewers into right-wing extremism while, separately, unwittingly catering to paedophiles by recommending children's videos. Grey zone examples (due to borderline legality depending on jurisdiction) include Airbnb devastating rental markets and exacerbating housing issues for low-income residents in various cities, as well as Uber negatively impacting public transit ridership while increasing vehicles on roads.

notion of “emergence” as applied to platform-facilitated systemic harm to marginalized communities; cases that turn out to involve only quasi- or pseudo-emergence can be disposed of with little difficulty under traditional notions of foreseeability, even with algorithms involved. However, this may require a recalibration of reasonable foreseeability. Part V provides this recalibration by demonstrating that reasonable foreseeability should inherently encompass *intersectional foreseeability* and *relational foreseeability* (based on Carys Craig’s relational copyright and Ian Kerr’s relational privacy). The intersectional foreseeability analysis draws on feminist tort theory and Canadian jurisprudence and legal scholarship advocating more contemporary interpretations of the reasonable person, whose hypothetical mindset is used to determine what is “objectively” reasonably foreseeable for the purposes of determining liability.

Part V addresses how the foreseeability analysis is impacted where a platform-facilitated systemic harm was truly emergent and genuinely unforeseeable, due to the involvement of unsupervised autonomous algorithms. Drawing on academic literature regarding robotics and AI liability, including in the contexts of autonomous vehicles and medicine, Part V proposes a tentative sliding scale of liability as a starting point for discussion, based heavily on work by Yavar Bathaee. Under the proposed framework in this paper, the standard of liability depends on the initially perceivable risk level of the impugned platform algorithm(s) and the level of harm that actually occurred in the case. The lowest level is a no-fault insurance pool for low- and medium-risk algorithms that resulted in low-level harm, then a modified negligence test for low- and medium-risk algorithms that resulted in medium- or high-level harm, and finally a strict liability standard for high-risk algorithms, regardless of the level of actual harm caused. The modified negligence test includes two key distinctions from the traditional negligence test, taken from Bathaee, and further considerations rooted in the context of platform-facilitated emergent systemic harms and the platform foreseeability gap, as developed throughout earlier sections of this paper.

Part I. Platform-Facilitated Emergent Systemic Harm

Part I of this paper will define what constitutes an emergent systemic harm arising from online platforms (used interchangeably with “digital platforms”). It will also discuss in more detail several factors that may increase the likelihood of a platform algorithm contributing or giving rise to one or more emergent systemic harms. In defining the characteristics and contributing factors of emergent systemic harm, the discussion will rely on real-world examples where harmful consequences have resulted from platform algorithms, in conjunction with platform companies’ business models, activities, users, or a combination of any or all of the above.

A. What Is Emergent Systemic Harm?

i. Defining “Systemic Harm”

The key aspect and driving force of emergent systemic harm is its focus on collective harm that transcends harm that occurs to any one user, as a result of activities on online platforms. This analytical shift from individualized harms to collective, systemic harm and structural analysis is a hallmark of critical feminist legal theory, as well as other areas of critical legal scholarship rooted in equality-seeking movements, such as critical race theory, post-colonial theory, and queer theory.[33]

[33] See e.g. Naomi R Cahn et al, "The Case of the Speluncean Explorers: Contemporary Proceedings" (1993) 61 Geo Wash L Rev 1754; and Bender, *supra* note 16.

Emergent systemic harm refers to harm that exists on a level beyond the harm experienced by users on an individual basis, such as the case of any one individual experiencing online harassment, targeted abuse, privacy violations, discriminatory exclusion, or voice silencing. This is not to say that online harassment, targeted abuse, privacy violations, discriminatory exclusion, or voice silencing, and other harms visited upon platform users are outside the scope of this paper or cannot constitute emergent systemic harm. What it means is that the relevant analytical focus of any of these or other given harms is not the harm experienced in an individual context or potential remedies in the case of an individual, but rather how such harms may arise systemically even in the absence of intent on the part of the platform companies that enable and facilitate them, and how to determine the appropriate extent (if any) of liability in such cases.

McLeod-Kilmurray describes how the law often "undervalues harms to women, and emphasizes individual and pecuniary harms while failing to 'see' collective and non-pecuniary harms".[34] Feminist analysis remedies this legal deficiency by "expos[ing] the law's resistance to recognizing that there are a variety of harms and to understanding that *some harms are collective and systemic*."[35] Her application of this insight to environmental law can extend to apply equally to online platforms and the harms they engender. Specifically, McLeod-Kilmurray notes how certain interpretations of liability for environmental harm "us[es] the private realm and private law to mask what are in fact public issues and to shield harm from redress".[36] One might say the same of the areas of law most commonly engaged in the context of online platforms, demonstrated by how users' privacy rights are often subordinated to the constraints of commercial contract law, where individual uninformed users are deemed to have acted in comparable legal and business sophistication to multinational technology companies.[37]

Similarly, how McLeod-Kilmurray describes the court's reasoning in *Hoffman v. Monsanto Canada Inc.*, 2007 SKCA 47—"trying to fit the systemic environmental contamination dispute into a private dispute between neighbouring farmers, leaving the corporate generators of the problem out of the picture"[38]—parallels platform-exonerating arguments regarding privacy violations, online abuse, and discrimination on the basis of protected traits. We have seen this with Facebook attempting to outsource privacy obligations to third-party app developers,[39] Twitter refusing to penalize or take responsibility for users engaging in online abuse,[40] and Google insisting that its search engine results only reflect, and do not shape, Internet users' beliefs and perceptions about the world.[41] In each of these and other similar cases, the wrongdoing is recast as attributable purely to private actors (individual app developers, individual harassers, individual online searchers, all acting in their own private capacities) who harmed individual private actors (victimized individual users, in their own private capacities),[42] in a way that invisibilizes

[34] McLeod-Kilmurray, "Ecofeminist Legal Critique", *supra* note 16 at 144-45.

[35] *Ibid* at 144 (emphasis added).

[36] *Ibid* at 143.

[37] Pavlovíc, "Contracting Out", *supra* note 14 at 421-22.

[38] McLeod-Kilmurray, "Ecofeminist Legal Critique", *supra* note 16 at 161.

[39] See e.g. See e.g. *Report of Findings into the Complaint Filed by the Canadian Internet Policy and Public Interest Clinic (CIPPIC) against Facebook Inc Under the Personal Information Protection and Electronic Documents Act by Elizabeth Denham Assistant Privacy Commissioner of Canada* (16 July 2009), PIPEDA Report of Findings #2009-008, online: Office of the Privacy Commissioner of Canada <www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2009/pipeda-2009-008/>; and *Joint investigation of Facebook*, *supra* note 28.

[40] Casey Newton, "Twitter's CEO keeps substituting talking for doing", *Verge* (24 January 2019), online: <www.theverge.com/2019/1/24/18195245/jack-dorsey-twitter-media-tour-2019> [Newton, "Twitter's CEO"].

[41] Noble, *supra* note 7 at 82.

[42] This may represent a twist on the "stupid user" doctrine, where blame instead falls on their evil twin, the malicious user. Jennifer Barrigar, "Submission: Office of the Privacy Commissioner of Canada Consultation – Online Reputation" (August 2016), online: <https://www.priv.gc.ca/en/about-the-opc/what-we-

each platform’s role as, if not corporate generators, certainly corporate enablers and corporate beneficiaries of the harm caused to and suffered by marginalized and vulnerable communities.

Part of applying a “broader view of what constitutes harm” includes critiquing the pervasive rhetoric of “choice”, by recognizing how what appears to be the freely made decisions of members of communities protected under non-discrimination laws, are “in fact restricted by external realities”:[43]

> The availability of any choice must be examined in its context which would include factors pertaining to race, class, gender, (dis)ability, age, or sexual orientation that may limit or coerce ‘choice’ or render a choice in effect unchoosable. …A contextualized approach to choice raises questions such as, choice to do what? What is being chosen between? And at what costs to whom? Who gets to decide what choices are made available? And on what basis are those decisions made? Who has choice? Who doesn’t? Why is this a choice to be protected?[44]

The context of online platforms compounds the restriction of choice applied to marginalized users. Their choices are first constrained by external forces in society generally speaking, as McLeod-Kilmurray indicates above. In addition to that, their choices as users of a particular platform are further constrained by that platform’s terms and conditions, technological affordances, design, and the behaviour of other users on the platform (such as those who would target marginalized users with online harassment). Additionally, Pavlovíc observes that these contractual abrogations of “choice” govern increasingly greater swaths of consumers’ lives:

> Business-to-consumer transactions in a globalized digital economy are governed almost exclusively by non-negotiable standard-form contracts, which are presented to consumers on a take-it-or-leave-it basis by all market players. Put differently, “[i]n the mass market, consumers are contract takers” and their access to and use of goods and services is conditional upon them accepting the terms of the standard-form contracts. The seemingly unlimited choices available to consumers are, in fact, constrained by the often onerous terms of standard-form contracts.[45]

What may first appear solely as harm to an individual user thus becomes, in reality, a symptom of layers of systemic forces coalescing into harmful patterns that disproportionately impact marginalized communities across the board (such as in instances where they have fewer alternative options to a given undesirable situation).

Overlooking systemic analysis while focusing on individual cases can also lead to the law inadequately recognizing liability even where it recognizes harm occurred. For example, in their paper comparing the legal liability regime surrounding driving offences to that surrounding safety offences in workplace environments, Bittle and Snider demonstrate how political and economic priorities and assumptions shape law regarding the latter in a way that mitigates findings of liability for corporate negligence.[46] The authors explain:

> [D]espite the considerable harms associated with both activities, safety crimes [by companies] generally are depicted in legal discourses as accidents, non-criminal breaches of particular regulations to be penalized through administrative penalties such as fines, while driving offences that lead to injury and/or property loss are seen as culpable, potentially criminal events best enforced

do/consultations/completed-consultations/consultation-on-online-reputation/submissions-received-for-the-consultation-on-online-reputation/or/sub_or_08/>.

[43] McLeod-Kilmurray, “Ecofeminist Legal Critique”, *supra* note 16 at 146.

[44] *Ibid* at 146 (footnotes omitted).

[45] Pavlovíc, “Contracting Out”, *supra* note 14 at 392 (footnotes omitted).

[46] See generally Steven Bittle & Laureen Snider, "Law, Regulation, and Safety Crime: Exploring the Boundaries of Criminalizing Powerful Corporate Actors" (2015) 3 CJLS 445.

> by police officers and penalized by everything from fines to licence suspension to incarceration. […]
>
> Our argument is that much of this complexity is legally and socially shaped by business interests, with the worldview or common sense of corporate capitalism significantly shaping the development and enforcement of new laws affecting corporations and corporate actors. Moreover, capitalist societies cling to the fiction that workers are free to choose their workplace and therefore voluntarily assume this risk. The traffic victim, conversely, is seen as having risk thrust upon him/her, she is an unwilling victim.[47]

In the context of online platforms, users are similarly "responsibilized"[48] for assuming the risks of using any particular platform, an agreement nominally enshrined in each platform's terms of service, or terms and conditions. Such users are treated in law as if they entered into contracts with each respective platform on the basis of equivalent freedom, sophistication, and negotiating power as the platforms themselves. Treating potential harms arising from platform algorithms as a matter of individual actions and harms alone untenably narrows the field of legal vision, as it is in the critical mass of actions and how the operate within the platform's business model and technological affordances as a whole that eventually gives rise to systemic harms as described in this paper. The harm itself is borne of the system as a whole, not of any individual unit within it. Focusing on individual users, rather than the aggregate consequences of their actions, causes one to miss the emergent systemic forest for the individually "innocuous" trees.

Bittle and Snider's further discuss opponents of corporate liability "individualiz[ing] guilt" among workers while simultaneously dismissing the notion that "a company could be held criminally liable if management allowed or encouraged a corporate culture that facilitated law violation or avoidance […] The belief that corporations are an unequivocal good, a vehicle for realizing our collective social and economic well-being, permeated the entire law-passage process."[49] Here we see parallels to the online platform context, where individual employees, users, or otherwise bad actors are scapegoated as the sole causes of particularly harmful consequences, while the role of the platform itself and how it may have actively facilitated such harms are smoothed over or waved away.[50] The notion of corporations being an "unequivocal good" recalls Shoshana Zuboff's writing regarding surveillance capitalism and how its purveyors operate by way of unchallenged "declarations" taken for granted as truth, such as the declaration that all data from all bodies and experiences is inherently fair game and up for grabs, and often for free.[51] The combination of the presumed good of the platform, with individually allocated responsibility for harm, produces and conceals analytical cracks into which platform-facilitated emergent systemic harms fall. Identifying, preventing, mitigating, or remedying systemic harm thus requires remaining alert to all systemic factors of an online platform, joint impacts from systemic factors working together or interacting in unexpected ways with potential collateral damage; and the potential collective

[47] *Ibid* at 447-48.

[48] "The Keynesian notion of the 1960s and 70s that the state is responsible for protecting workers against bad employers has been replaced by neoliberal discourses that 'responsibilize' workers, treating them as 'partners' with management with mutual goals and equal responsibility for worker safety (Tombs and Whyte 2007)." *Ibid* at 452.

[49] *Ibid* at 450.

[50] For an "offline" example of this, see Google paying a former executive to quietly resign in response to sexual assault allegations, rather than address systemic issues within the company that saw sexual harassment and sexual assault treated leniently, until its own employees pressured the company to implement change. "Google paid former executive $35m after sexual assault allegation", *Guardian* (11 March 2019), online: <www.theguardian.com /technology/2019/mar/11/google-executive-payout-harassment-amit-singhal>; Matthew Weaver et al, "Google walkout: global protests after sexual misconduct allegations", *Guardian* (1 November 2018), online: <https://www.theguardian.com/technology/2018/nov/01/google-walkout-global-protests-employees-sexual-harassment-scandals>.

[51] Zuboff, *supra* note 5 at 177 and 193-93 (regarding "declarations") and 210-11 (regarding "dark data").

impact of any given user actions writ large, multiplied across a platform's entire user base, particularly if the impact disproportionately affects members of historically marginalized groups.

As a qualification, this paper focuses on platform-facilitated emergent systemic harm to marginalized groups specifically, in the context of equality and freedom from discrimination. However, this is not the only form of platform-facilitated emergent systemic harm. Systemic harm can also occur in the context of competition law (harm to the state of market competition), or in the context of democratic institutions (harm to political processes and electoral integrity through filter bubbles and the erosion of a certain level of public political discourse). Due to time and space constraints, however, these other forms of platforms-related systemic harms are excluded from the scope of this paper and could be investigated and analyzed by others through the lens of platform-facilitated emergent systemic harm in future research.

ii. Defining "Emergence"

The term "emergent" in emergent systemic harm has a specific meaning taken from the context of artificial intelligence and robotics as well as other fields that involve the study of complex systems, whether in physics, sociology, or environmental studies.[52] In the AI and robotics context, Calo explains emergence as "the ability or tendency of a system to behave in complex, unanticipated ways."[53] He builds on the work of Stephen Johnson, for whom emergence is "the movement of low-level rules to tasks of apparently high sophistication."[54] While Calo discusses emergence as a property of robots specifically, this paper and the notion of emergent systemic harm arising from online platforms is explicitly rooted in recognizing that "[e]ven absent embodiment an emergent system can threaten critical aspects of society, as when high-speed trading algorithms destabilize the stock market"[55]—or when recommendation algorithms turn the world's most popular user-generated video platform into a "radicalization engine" systematically driving users towards ever more extremist views[56]—without having been instructed or designed to achieve any such aim, *per se*. In short, emergence is when a system produces results that are more than the sum of its parts.[57] Emergent systemic harm, then, is systemic harm that unexpectedly results from the constituent parts of a complex system working together in a way that transcends what they and the system as a whole was intended and knowingly designed to do.

[52] "Emergence is not only common in physics but also in many other disciplines, including biological science where defining and delimiting ecosystems require the identification of emergent properties." See generally Hugo Tremblay, "Sustaining Development in a Thermodynamic Universe: Raging Against the Dying of the Light" (2016) 28:3 J Envtl L & Prac 333; see also hydrogen atoms example in *ibid.*, at 351 and biological organs example in Éric Labelle Eastaugh, "The concept of a linguistic community" (2018) 69:1 UTLJ 117 at 129.

[53] Calo, "Robots", *supra* note 18 at 40. Elsewhere, Calo refers to emergence as "unpredictably useful behavior": Calo, "Lessons", *supra* note 18 at 532. For the purpose of defining emergent systemic harm, "useful" should be understood to mean "functional" or "effective" in that the behaviour achieves some end, where in this context that end is causing harm of some sort. See Tufekci's notion of algorithms of "actants", "computational agents that are not alive, but that act with agency in the world", which she terms computational agency. Tufekci, "Computational Agency", *supra* note 6 at 207.

[54] Calo, "Lessons", *supra* note 18 at 539. Calo provides as an example "the way ants follow simple rules to accomplish complex, seemingly intelligent tasks". Another example, particularly pertinent to this paper, would be the way individual users reading and sharing articles—in itself a simple task—gave rise to unintended systemic consequences that could have been considered sophisticated if accomplished deliberately, such as fragmenting public discourse through social media filter bubbles, and kicking the legs out from underneath traditional journalism and media. See e.g. Robert W McChesney, *Digital Disconnect: How Capitalism is Turning the Internet Against Democracy* (New York: The New Press, 2013).

[55] Calo, "Robots", *supra* note 18 at 40-41.

[56] Tufekci, "Great Radicalizer", *supra* note 6.

[57] Tremblay, *supra* note 52 at 351.

B. Constituent Components of Platform-Facilitated Emergent Systemic Harms

The constituent parts of a complex system contributing to emergent systemic harm include: the platform's algorithms relevant to the features implicated in harmful consequences (whether identified as such before or after the fact); the online platform's business model; the technological affordances that facilitate what users are able to do on a given platform (whether the capability was intended by the platforms' owners and designers or not); the platform's scale as indicated by reach and number of users; the passage of time (generally, but not always, a function of scale); and users' views and usage of platforms as individualistic and instrumental (as explained by Gillespie).

The following subsections will examine each of the above contributing factors in turn. Note that not all of the examples described below will themselves fit the definition of "emergent systemic harm", but rather are provided to illustrate the specific factor under discussion. For instance, cases involving technology-facilitated gender-based violence are provided to illustrate the concept of technological affordance. Such violence is an example of a systemic harm that is not emergent because it involved bad actors intentionally harming others, and in a way that was not only foreseeable, but many would consider predictable if not all but certain. Examples of emergent systemic harm that bring together all of the contributing factors discussed will be provided at the end of this section.

i. The Platform's Algorithms

The first constituent element of emergent systemic harm is a platform's algorithms. An algorithm is an automated computational "set of instructions for carrying out procedures", and the term "platform algorithms" refers to "computational processes that are used to make decisions of such complexity that inputs and outputs are neither transparent nor obvious to the casual human observer", in the context of online platforms. [58] An algorithm may be optimized to achieve one particular objective, that itself was legitimate and reasonable, but end up achieving it in an unexpected manner which causes collateral consequences, or end up achieving a completely different objective instead, due to mismatch between the coders' intentions and how the algorithm operates in practice, or in conjunction with all of the other contributing factors of platform-facilitated emergent systemic harm. At the same time, algorithmic decision-making, or what Tufekci calls "computational agency",[59] is "expanding into more and more spheres. Complex, opaque and proprietary algorithms are increasingly being deployed in many areas of life, often to make decisions that are subjective in nature".[60] The complex, opaque, proprietary, and subjective[61] nature of algorithms makes for a perfect storm of potential harm, both in the context of online platforms and in real-life consequences beyond the platforms themselves. A significant and increasingly high-profile field of scholarship and research has emerged to address such issues in recent years, broadly grouped under terms such as algorithmic accountability, algorithmic transparency, AI ethics, or AI/algorithmic fairness, accountability, transparency, and ethics (FAT(E)).[62]

Algorithms encompass multiple layers of emergence. First, they contribute to emergent systemic harm as set out in the definition, as constituent elements of larger complex systems that combine to engender results greater than the sum of their parts. Even before that, however, algorithms themselves are

[58] Tufekci, "Computational Agency", *supra* note 6 at 206.
[59] *Ibid* at 217.
[60] *Ibid* at 217.
[61] The use of algorithms in subjective decision-making means there are "no anchors or correct answers to check with [for instance, there is no such thing as a mathematically correct newsfeed]. Lack of external anchors in the form of agreed-upon 'right' answers makes their deployment especially fraught." Tufekci, "Computational Agency", *supra* note 6 at 217.
[62] See note 21, *infra*; see also "Fairness, Accountability, and Transparency in Machine Learning" (2018), online: <www.fatml.org> and "ACM FAT* Conference 2020", online: <fatconference.org/2020/>.

instruments of emergence, particularly unsupervised machine learning algorithms that are specifically designed and used precisely for their ability to work in ways that humans would not predict or be able to conceptualize or accomplish. For reasons such as the algorithm's design, poor training data, inaccurate input data, or the inner workings of an algorithm unintelligible to human understanding, algorithms can and have routinely resulted in "emergent bias", or bias that "emerge[s] unexpectedly from decisional rules developed by the algorithm, rather than any 'hand-written' decision-making structure".[63] This is on top of an algorithm's (or separately, a platform's) non-emergent technical bias, which "arises from technological constraints, errors or design decisions, which favour particular groups without an underlying driving value".[64] Thus, platforms' increasing reliance on algorithms and "algorithmic gatekeeping"[65] warrants particularly attention when evaluating a platform or feature for potential emergent systemic harms.

Furthermore, Tomer Shadmy and Danny Butt have both analyzed how platform algorithms increasingly engage in mass governance of populations in ways that bypass traditional democratic processes. This alone makes the use of algorithms a *systemic* issue, and not one that can be adequately analyzed or addressed by narrowing one's viewing window to that of any single user. For one thing, "[a]s algorithms spread globally and consolidate their rearchitecting power, their archival authority is private and immune to the user, reserved for the platform architects who decide from afar which modes of informational behaviour are most profitable."[66] Shadmy also explains how machine-learning algorithms constrain users' choices on platforms—often in a way designed to maximize monetization and profitability—and subtly yet systematically shape users' environments and their understanding of such, and thus their respective options and eventual actions:

> These machine-learning systems not only predict the choices of the users based on their behavior but also create their choices in some senses. The algorithmic analysis of data patterns dynamically configures the targeted individual's choice environment in highly personalized ways, affecting individuals' behavior and perceptions by subtly molding the networked user's understanding of the world that surrounds them. … For example, if, for one reason or another, a user is deemed compatible with the profile of people who buy one brand of shoes, Facebook will expose him or her to these shoe-related advertisements and content. This raises the chances that the user will also become interested in buying the shoes, even if he originally had no particular interest in that type of product. Hence, the user's "choice" is created by the environment that the platform produces. The platform's algorithms imagine the users, and they respond accordingly within the algorithms' affordances. Via continuous feedback loops based on online users' interactions, algorithms configure individuals online by "tailoring their conditions of possibility". [67]

Viewed through this lens, algorithms might be considered to be the heart of emergent systemic harm that arises from online platforms. Algorithms are by no means the only contributing factor, by definition, but

[63] Mittelstadt et al, *supra* note 17 at 8.

[64] "Examples include when an alphabetical listing of airline companies leads to increase business for those earlier in the alphabet, or an error in the design of a random number generator that causes particular numbers to be favoured. Errors can similarly manifest in the datasets processed by algorithms. Flaws in the data are inadvertently adopted by the algorithm and hidden in outputs and models produced." *Ibid* at 7 (in-line citations omitted).

[65] "Algorithmic gatekeeping is the process by which such non-transparent algorithmic computational-tools dynamically filter, highlight, suppress, or otherwise play an editorial role—fully or partially—in determining: information flows through online platforms and similar media; human-resources processes (such as hiring and firing); flag potential terrorists; and more." Tufekci, "Computational Agency", *supra* note 6 at 207-08.

[66] Danny Butt, "New International Information Order (NIIO) Revisited: Global Algorithmic Governance and Neocolonialism" (2016) 27: *Fibreculture J*, online: <twentyseven.fibreculturejournal.org/2016/03/08/fcj-198-new-international-information-order-niio-revisited-global-algorithmic-governance-and-neocolonialism/> at 32.

[67] Tomer Shadmy, "The New Social Contract: Facebook's Community and Our Rights" 37:2 BU ILJ 307 at 348-49 (emphasis in original, footnotes omitted).

their involvement also significantly raises the likelihood that emergent systemic harm will result from a particular platform.

For example, Tufekci revealed that during the 2014 protests against police brutality and anti-black racism in Ferguson, Missouri, there was almost no mention of the ongoing events on Facebook, despite occupying a central role in people's timelines on Twitter:

> Facebook's algorithm had "decided" that such stories did not meet its criteria for "relevance"—an opaque, proprietary formula that changes every week, and which can cause huge shifts in news traffic, making or breaking the success and promulgation of particular stories or even affecting whole media outlets. By contrast, Twitter's algorithmically unfiltered feed allowed the emergence of millions of tweets from concerned citizens, which then brought the spotlight of the national media. Algorithmic filtering also by Twitter might have meant that a conversation about police accountability and race relations that has since shaken the country might never have made it out of Ferguson. […]
>
> Given that so many protests and social movements depend on new media, especially to circumvent censorship and to organize, it is especially important to pay attention to the role algorithmic connectivity plays in civic ecology.[68]

Platform curation algorithms have also influenced political issues on a more longterm and fundamental basis. One of the most prominent examples of algorithms going awry on a social media platform, in this context, is YouTube's recommendation algorithm. Over the past several years and around the world, academics have documented how YouTube has become an accelerator of political extremism and right-wing radicalization, through its recommendation engine and autoplay features combining to create a "rabbit hole" of continuous viewing, even while the company denied both the phenomenon and any intention to create it.[69] It is important to note that YouTube creating algorithms that would recommend videos to users, based on likelihood of viewer "engagement", is on its face a legitimate and reasonable thing for a company to do. However, researchers have found on repeated occasions and in separate, independent studies, that "YouTube's search and recommendation system appears to have systematically diverted users to far-right and conspiracy channels" in Brazil,[70] the United States, Canada, and

[68] Tufekci, "Computational Agency", *supra* note 6 at at 213 and 215 (note as well the mention of the role of algorithms as part of an "ecology", or interconnected system). Tufekci also notes, "Later, I performed a Twitter search for the keywords 'Twitter Facebook Ferguson' and found that hundreds of ordinary people were complaining of a similar information blackout on Facebook, that was instead dominated by the 'ice bucket challenge' in which people poured buckets of ice water on themselves in support of a charity, and invited their network to the same." *Ibid* at 214. See also Carlos Lozada, "Twitter and Facebook help spark protest movements. Then they undermine them", *Washington Post* (25 May 2017), online: <https://www.washingtonpost.com/news/book-party/wp/2017/05/25/twitter-and-facebook-help-spark-protest-movements-then-they-undermine-them/>.

[69] See e.g. "YouTube says there is no rabbit hole effect. 'It's not clear to us that necessarily our recommendation engine takes you in one direction or another,' said Ms. O'Connor, the product director." Max Fisher & Amanda Taub, "On YouTube's Digital Playground, an Open Gate for Pedophiles", *New York Times* (3 June 2019), online: <www.nytimes.com/2019/06/03/world/americas/youtube-pedophiles.html> [Fisher & Taub, "Open Gate"]. This quote was provided in the context of another systemic harm that emerged from YouTube, that of catering to paedophiles with children's videos; however, nearly all of the platform-side factors remain the same, particularly the role of YouTube's recommendation algorithms.

[70] Max Fisher & Amanda Taub, "How YouTube Radicalized Brazil", *New York Times* (11 August 2019), online: <www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html> [Fisher & Taub, "Brazil"]. The article also states, "By repeating this thousands of times, the researchers tracked how the platform moved users from one video to the next. They found that after users watched a video about politics or even entertainment, YouTube's recommendations often favored right-wing, conspiracy-filled channels like Mr. Moura's. Crucially, users who watched one far-right channel would often be shown many more. *The algorithm had united once-marginal channels — and then built an audience for them*, the researchers concluded" (emphasis added).

Germany.[71] Moreover, YouTube's algorithms did not simply direct users to a pre-existing networked web of right-wing extremism and conspiracy theories; they played a central role in *creating* that networked universe, [72] by connecting and promoting to viewers channels that would have otherwise likely remained isolated and marginal.[73] In fact, Tufekci observed the following while also noting the unintended nature of the harm and its resulting from multiple forces, each "innocent" in itself, combining to bring about terrible consequences:

> 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.
>
> This is not because a cabal of YouTube engineers is plotting to drive the world off a cliff. A more likely explanation has to do with the nexus of artificial intelligence and Google's business model. (YouTube is owned by Google.) For all its lofty rhetoric, Google is an advertising broker, selling our attention to companies that will pay for it. The longer people stay on YouTube, the more money Google makes.[74]

Guillaume Chaslot, who worked on YouTube's recommendation algorithms while a computer programmer at Google, began warning the public about the platform's effects based on his experiences as an employee and on his own subsequent research into how YouTube's algorithms operated on users in practice.[75] He concluded, as has become clear to many by now, "The recommendation algorithm is not optimising for what is truthful, or balanced, or healthy for democracy."[76] This is despite the fact that those who created, work at, manage, and own YouTube and Google presumably did not intend to optimize their powerful platforms to work against truth, balance, or democracy around the world.

ii. The Platform's Design and Technological Affordances

The second factor that is likely to contribute to emergent systemic harms from online platforms is their technological affordances, or what user actions the platforms enable, encourage, or discourage by how they are designed. The term "affordance" is itself the focus of a body of literature in its own right, and one whose use, meaning, and definitions have shifted and migrated across different fields, from its origins in

[71] See e.g. Caroline O'Donovan, "We Followed YouTube's Recommendation Algorithm Down The Rabbit Hole", *BuzzFeed* (24 January 2019), online: < https://www.buzzfeednews.com/article/carolineodonovan/down-youtubes-recommendation-rabbithole>; Kevin Roose, "The Making of a YouTube Radical", *New York Times* (8 June 2019), online: <www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html?module=inline>; Paul Lewis, "'Fiction is outperforming reality': how YouTube's algorithm distorts truth", *Guardian* (2 February 2018), online: <www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth>; Max Fisher & Katrin Bennhold, "As Germans Seek News, YouTube Delivers Far-Right Tirades Image", *New York Times* (7 September 2018), online: <www.nytimes.com/2018/09/07/world/europe/youtube-far-right-extremism.html>; Jonas Kaiser & Adrian Rauchfleisch, "Unite the Right? How YouTube's Recommendation Algorithm Connects The U.S. Far-Right" (11 April 2018), online: *Medium* <medium.com/@MediaManipulation/unite-the-right-how-youtubes-recommendation-algorithm-connects-the-u-s-far-right-9f1387ccfabd>.

[72] "In this blogpost, we share some preliminary results of our analysis of 13,529 channels, which shows how YouTube's recommendation algorithm contributes to the formation of a far-right filter bubble. […] In our analysis, we show that YouTube's recommendation algorithms actively contribute to the rise and unification of the far-right." Kaiser & Rauchfleisch, *supra* note 71.

[73] "As our data shows, the channel recommendation connects diverse channels that might be more isolated without the influence of the algorithm, and thus helps to unite the right." *Ibid*.

[74] Tufekci, "Great Radicalizer", *supra* note 6.

[75] Paul Lewis & Erin McCormick, "How an ex-YouTube insider investigated its secret algorithm", *Guardian* (2 February 2018), online: <www.theguardian.com/technology/2018/feb/02/youtube-algorithm-election-clinton-trump-guillaume-chaslot>.

[76] Lewis, *supra* note 71.

environmental psychology, to media studies, to industrial (physical) and graphic (including web) design, to human-computer interactions, and to science and technology studies.[77]

One of the major distinctions between different definitions of an affordance turns on whether or not user perception is required for something to constitute an affordance. For example, according to Dan Norman, a possible action—such as clicking on a link—is not an affordance unless the user *perceives* that the action is possible. According to James Gibson, a possible action is an affordance regardless of whether the user realizes it or not. To illustrate, imagine a perfectly functional hyperlink that is technically clickable, but does not look any different from the surrounding body text on a website, nor does the cursor change when it hovers over the link. According to Gibson, the link is an affordance because it still works and can be clicked, even if the user does not perceive its "clickability". According to Norman, there is no affordance because nothing signals that any possible action of clicking is open to the user, unless they actually click on it.[78]

For the purposes of this paper and in the context of online platforms, an affordance—used interchangeably with technological affordance, or platform affordance—is defined according to Gibson: any possible action that a user can take on an online platform is an affordance, even if the user does not realize they can take that action. In part this is because a user perception-dependent definition of affordance would become highly subjective and shift with every user: under Norman's definition, any possible action on a platform is simultaneously an affordance for some users while not being an affordance for other users. However, as will become clear shortly, Norman's definition is crucial to understanding the contribution of platform affordances to emergent systemic harm.

Platform affordances contribute to emergent systemic harms in two overarching ways. The first way involves affordances that are essentially universally perceived both by the platform company (specifically, its engineers, designers, managers, and owners) and by the platform's users. These affordances are likely what the platform promotes, and what users perceive, as features: actions that are possible to take on the platform, which the platform company intends users to take, and which users perceive they are invited to take and do take. Emergent systemic harm arises from universally perceived affordances where they have unintended consequences.

An example of one emergent systemic harm that emerged in part from such affordances would be the breakdown of a shared factual backdrop for public political discourse in countries where Facebook is a dominant social media network. The affordances in this case are the ability to curate one's own newsfeed (prioritizing certain people and pages while excluding others), the ability to "like" or otherwise "react" to others' posts, and the ability to share links with one's own social network. All of these actions are neither illegal nor unethical, and in fact are perfectly reasonable things for an individual to do in the course of going about their online activities and shaping their preferred online experience. However, a critical mass of users engaging in these activities, combined with other factors discussed in this section—such as an ad-driven business model and what the platform's algorithms are optimized to do—results in fragmentation and digital balkanization of public discourse, where every individual user has their own version of reality

[77] See e.g. Donald Arthur Norman "Affordance, conventions, and design" (1999) 6:3 interactions 38; Joanna McGrenere & Wayne Ho, "Affordances: Clarifying and Evolving a Concept" (Paper published in the Proceedings of Graphics Interface 2000, Montreal, May 2000), online (pdf): <www.cs.ubc.ca/~joanna/papers/GI2000_McGrenere_Affordances.pdf>; H Rex Hartson "Cognitive, physical, sensory, and functional affordances in interaction design" (2003) 22:5 Behaviour & Inf Tech 315; and Jeroen Stragier et al, "Understanding persistence in the use of Online Fitness Communities: Comparing novice and experienced users" (2016) 64 Computers in Human Behavior 34 at 35.

[78] This explanation is necessarily simplified, and other scholars have made further distinctions. For example, McGrenere and Ho distinguish between perceptible affordances and hidden affordances, while Hartson distinguishes between cognitive affordances, physical affordances, sensory affordances, and functional affordances. *Ibid.*

and public opinion as represented by their Facebook newsfeed, or their own version of a public water cooler as represented by their Twitter timeline. However, they mistake these personalized representations for "*the*" public forum or "*the*" water cooler, the one everyone else also sees, not just one of several billion versions specifically tailored to them based on earlier engagement with others' content, the links they shared, and the way they curated their own feed.

The second way platform affordances contribute to systemic emergent harms is through *differentially perceived* affordance: harm arises from the very fact that only certain actors perceive the affordance, while others do not realize it is there at all, or did not intend the affordance to be discovered or used. It is here that Norman's definition becomes relevant: where a user perceives possible actions that others don't, the platform *has more affordances* for them than for all other users on the same site, giving that user an advantage over other users who don't have that affordance because they did not perceive the possibility.

This gap between platform affordances existing for users who perceive it and those who do not most clearly arises with the issue of online harassment, targeted abuse, and technology-facilitated gender-based, race-based, sexual orientation-based, and other protected characteristics-based abuse and violence on social media platforms. Systemic harm arises as a result of platform affordances that enable abuse being perceived by both abusers and victims, but ostensibly were not perceived or were dismissed by the platform company itself while designing the platform or designing the feature that enabled abuse. For example, in 2017, Snapchat released a feature, "Snap Map", that allowed users to track each other's precise location in real time, with users' names and profile pictures plotted on a map with pinpoint accuracy. While users had the option to turn this feature off, Snapchat did not make clear to users that the app would also publish their exact location to their entire friends list every time the user opened the app.[79] Privacy advocates and those concerned for children's safety as well as violence against women immediately raised concerns that such a feature would easily allow stalking, harassment, and privacy violations of unsuspecting users, particularly women or survivors of intimate partner violence.[80] In fact, the app "quickly became a tool … to spot cheaters", joining the ranks of other types of technology tied to "catch cheating spouses" rhetoric and which are closely associated with intimate partner violence and abuse, such as stalkerware.[81] In one case, a man stalked his girlfriend via Snap Map, located her in person,

[79] Dani Deahl, "Snapchat's newest feature is also its biggest privacy threat", *Verge* (23 June 2017), online: <www.theverge.com/2017/6/23/15864552/snapchat-snap-map-privacy-threat>.

[80] See e.g. Olivia Solon, "Snapchat's new map feature raises fears of stalking and bullying", *Guardian* (23 June 2017), online: <www.theguardian.com/technology/2017/jun/23/snapchat-maps-privacy-safety-concerns>; Corbin Streett, "What's the Deal with Snap Map?" (6 June 2017), online: *Technology Safety* <www.techsafety.org/blog/2017/7/6/whats-the-deal-with-snap-map>; JR Thorpe, "Will Turning On Snap Map Put You At Risk?" *Bustle* (10 July 2017), online: <www.bustle.com/p/will-turning-on-snap-map-put-you-at-risk-67358>; Arianna Chatzidakis, "Snapchat's 'Snap Map' Is A Major Privacy Threat And We're So Not Here For It", *Grazia* (9 July 2017), online: <graziadaily.co.uk/life/opinion/snapchat-snap-map-ghost-mode/>; Maggie Nicholson, "Snapchat's new Snap Map feature may pose a threat to victims and survivors" (19 July 2017), online: *National Latin@ Network* <enblog.nationallatinonetwork.org/snapchats-new-snap-map-feature-may-pose-a-threat-to-victims-and-survivors/>.

[81] "Spyware and other forms of technology that facilitate [intimate partner surveillance] are sometimes referred to as stalkerware. In some circumstances, stalkerware technology is used in an intimate relationship to conduct powerfully intrusive covert or coerced surveillance of an intimate or former partner's mobile device without their knowledge. Once installed, stalkerware apps allow an operator to access an array of intimately personal information about the surveillance target. The apps can enable realtime and remote access to text messages, emails, photos, videos, incoming and outgoing phone calls, GPS location, banking or other account passwords, social media accounts, and more. Stalkerware apps are sometimes used covertly while, in other circumstances, the technology is used openly to intimidate, harass, or extort the surveillance target." Cynthia Khoo, Kate Robertson & Ronald Deibert, "Installing Fear: A Canadian Legal and Policy Analysis of Using, Developing, and Selling Smartphone Spyware and Stalkerware Applications" (June 2019), online (pdf): Citizen Lab <citizenlab.ca/docs/stalkerware-legal.pdf> at 1.

and stabbed another man who was with her in a car.[82] Snap, the company owning Snapchat, either underappreciated or did not perceive[83] Snap Map's affordances of stalking, violence against women, harm to children, or intimate partner abuse, while these affordances were instantaneously clear to both abusers and potential victims, or those who belong to groups disproportionately targeted by technology-facilitated violence, such as women and girls. The question, of course, is why such affordances were not brought to the attention of Snap Map's creators, developers, and management before the feature was unrolled—or why relevant concerns were not adequately addressed, or the full capabilities of Snap Map made clear to users upon roll-out to mitigate such harms or allow users to make meaningfully informed decisions.

Another example of a platform affordance that was perceived by users, but was not adequately perceived by the creators and owners of the platform itself, is the affordance of waging coordinated abuse, silencing, and account banning campaigns on members of marginalized communities on Facebook, through its content moderation flagging feature. Crawford and Gillespie provide a fulsome account of how users have "gamed" Facebook's flagging feature, turning it from its intended role as an "individualized mechanism of complaint"[84] to a variety of other "social and tactical" functions, including "user-generated warfare" on marginalized platform users:[85]

> In 2012, accusations swirled around a conservative group called "Truth4Time," believed to be coordinating its prominent membership to flag pro-gay groups on Facebook. One of the group's administrators claimed that this accusation was untrue, and that the group had formed in response to pro-gay activists flagging their antigay posts. Either way, it seems that surreptitious, organized flagging occurred. […]
>
> Strategic flagging is most prominent when visibility is perceived to be a proxy for legitimacy. As Fiore-Silfvast (2012) describes, a group of bloggers angered by the presence of pro-Muslim content on YouTube began an effort called "Operation Smackdown." Launched in 2007 and active as recently as 2011, the group coordinated their supporters to flag specific YouTube videos under the category of "promotes terrorism"…. [86]

Recognizing differentially perceived affordances is key to parsing the breakdown between a platform company and its users in the event of emergent systemic harm. Strategic and coordinated collective use of Facebook's flagging feature is considered "gaming" only because Facebook did not *intend* for it to be used that way. This lack of intention was not built into the technology itself, however. From a technical perspective, coordinated flagging campaigns were not gaming but a perfectly legitimate use of the tool—the fact it *could* be used that way meant, simply, that it could be used that way. The platform afforded it.

[82] Ms Smith, "Guy tracks down girlfriend via Snapchat Snap Map, stabs man she's with" (7 November 2017), online: CSO <www.csoonline.com/article/3236486/guy-tracks-down-girlfriend-via-snapchat-snap-map-stabs-man-shes-with.html>.

[83] "The safety of our community is very important to us," said a Snapchat spokesman, who added that location-sharing is off by default and "completely optional". "Snapchatters can choose exactly who they want to share their location with, if at all, and can change that setting at any time. It's also not possible to share your location with someone who isn't already your friend on Snapchat, and the majority of interactions on Snapchat take place between close friends." Solon, *supra* note 73. This statement appears not to appreciate that intimate partner violence, abuse, and harassment, by definition, takes place between individuals who are or were once close to each other, and that victims and survivors may not be in a position to digitally "defriend" their abuser, without facing violent repercussions in real life.

[84] Kate Crawford & Tarleton Gillespie, "What is a flag for? Social media reporting tools and the vocabulary of complaint" (2016) 18:3 new media & society 410 at 420.

[85] Gillespie, *Custodians*, *supra* note 10 at 420-421.

[86] Crawford & Gillespie, *supra* note 84 at 421.

Gillespie makes this point even more clearly with an example of one individual who took it upon himself to get drag queens suspended *en masse* from Facebook.[87] The user, known as @RealNamePolice, had reported hundreds of drag queens' accounts as fake and in violation of Facebook's "real name" policy, due to their use of stage names, deliberately targeting them based on biblical beliefs expressed through his Twitter and Tumblr accounts.[88] Although Facebook apologized for the mass account suspensions and modified their policy to allow for drag queens' stage names, Gillespie points out:

> [T]he paradox here is that while @RealNamePolice's motivations may have been political, and to some reprehensible, he did flag "appropriately": he did understand the policy correctly, and he did identify names that violated it. Was this a misuse of the flagging system, then, or exactly what it was designed for? Is flagging [meant to map user complaints, or deputize users to act on platform violations]? … Either way, handing the tools of policing to the community opens them to a variety of "uses" that move well beyond the kind of purpose-consistent efforts to "tidy up" that a platform might want to undertake, under a more editorial approach.[89]

What the results of these differentially perceived affordances suggest is that to meaningfully address or mitigate instances of systemic harm, as well as of emergent systemic harm, on their platforms, platform companies—including their employees, developers, engineers, designers, and management—must put greater effort into perceiving the range of affordances that their users clearly have no trouble identifying. At that point, they must either take ownership of the affordances or take the necessary steps to limit or remove their availability to users.

One more example of differentially perceived affordances leading to systemic harm flips the above dynamic: harm arises where platform affordances are known to the company, but not perceived by the users. Perhaps the most well-known example of this is online platforms' data collection and sharing practices: the user data collection, access, sale, aggregation, and inference capabilities on a platform are affordances that are clear to platforms and third parties such as their commercial partners, app developers, or advertisers, but are not always clear to the platform's users, and thus cannot inform their sharing and engagement decisions on the platform. Shoshana Zuboff's tome explores this particular set of platform affordances in-depth, as part of the underlying infrastructure and driving force behind what she describes as the rising era of surveillance capitalism.[90] Tomer Shadmy describes this dynamic as follows:

> Different apps, such as location and navigation apps, directly convey information on a user's whereabouts to the company and to their friends, without the user's direct and conscious action. Some future technologies could lead to a new scale of automation that further undermines the possibility of conscious choice—such as the technology aims to decode neural activity devoted to speech in the user's brain, and enable him or her to "type from the brain", directly, without a keyboard. Through these systems, deciding what thoughts, decisions, and sensations will be publicly visible will be an operation performed automatically by the company's technology.[91]

Shadmy also discusses another platform affordance that may lead to emergent systemic harm, that of Facebook's "reaction emojis" allowing users to react with one of exactly six crude emotions to the wide variety of content they encounter in their newsfeeds. Again, there is nothing inherently wrong with emojis or encouraging users to express emotions in response to a post. However, Shadmy explains how this feature, when used by the population at scale and in conjunction with the technological constraints on

[87] Gillespie, *Custodians*, *supra* note 10 at 94.
[88] *Ibid* at 94-95.
[89] *Ibid* at 95.
[90] See generally Zuboff, *supra* note 5. The most high-profile example of the systemic harm that arose from this set of affordances is, of course, the exploitation of users' personal data by Cambridge Analytica to leverage psychometric profiling in order to influence the 2016 United States presidential election, and the United Kingdom's referendum on whether or not to remain in the European Union. *Joint investigation of Facebook*, *supra* note 28.
[91] Shadmy, *supra* note 67 at 349. (footnotes omitted).

choice that both she and Zuboff explicate, may lead to "a new reliance on the freedom to feel, rather than freedom to choose, as the ultimate expression of individuality and therefore of rights. […] This ability and the right arising from it, the right to feel, one could say is far less threatening to Facebook's business model than other modern rights—such as the right to privacy."[92] What at first appears to be a harmless platform affordance—reacting to posts with a limited set of emojis—emerges as a subtle nudge shifting users from "classical" rights such as autonomy and privacy to a platform-mediated, stunted "right to feel", moreover one that erodes nuance[93] and the quality of online, and thus offline, public discourse over time.

For all of the examples above, the differential perception of whether or not an affordance exists goes directly to assessing to what extent a systemic harm may be considered "emergent". In the case of technology-facilitated gender-based abuse, such as with Snapchat's Snap Map, the harm to women and girls was "emergent" only to those who were ignorant of gender equality issues in society and the way that technology and online platforms intersect with such issues. To everyone else, the harm was to be expected from the moment of merely mentioning the term "location-tracking feature". It is through such scrutiny and parsing of who knew what when, and who neglected whom and what considerations, that the idea of "unintended consequences" from online platforms begins to crumble.

As another example, the systemic harm of ubiquitous privacy violations, the erosion of consent with respect to one's personal data, and the gradual co-optation of human autonomy[94] resulting from surveillance capitalism may be "emergent" to some while not to others. This harm would be "emergent" for the vast majority of users and non-users who do not keep abreast of online privacy and data protection developments, and did not realize the longterm, systemic impact of their collective online activities on data-collecting platforms. The systemic harm may also be "emergent" to unwitting technology companies (i.e., their owners, managers, and staff) who jumped on the ad-driven business bandwagon without considering, realizing, or being informed of the longterm implications for humanity more broadly. However, the systemic harms that surveillance capitalism introduces, which Zuboff details in her book, would not be considered "emergent" to those who were the architects of such a shift in business and society and did know what they were doing from the start, as Zuboff describes Google's founders and executives to have.[95] The systemic harms also would not have been "emergent" to Zuboff herself, or other academics who foresaw them early on. In fact, as Hugo Tremblay points out in the environmental law context, the idea of emergence is in some ways tautological:

> To some extent, the concept of emergence is problematic. Better knowledge about the constituent parts of a system might explain emergent properties. In that sense, the concept of emergence could be tautological. Yet, it remains useful given the impossibility of perfect knowledge. Emergence helps identify the critical thresholds in human development that should not be crossed for development to remain sustainable.[96]

The differing levels of knowledge Tremblay refers to in determining how "emergent" a particular consequence is directly relate to the differential perception of affordances on a digital platform, combined with broader intersectional awareness of history and systemic oppression in society, and how new technologies may manifest pre-existing power dynamics and abuses in new ways. The significance of this differential aspect of both platform affordances and the concept of emergence as applied to platform-facilitated systemic harms should inform future analysis regarding attributing potential liability to platforms for their role in facilitating systemic harms to marginalized communities, by way of actual or

[92] *Ibid* at 44-45.

[93] "According to this mindset and regulation by design, individuals should react to contents from a repertoire of six basic emotions, not with thoughts, logic, or opinions." *Ibid* at 44.

[94] See also Mireille Hildebrandt, "Profiling and the Rule of Law" (2008) 1:1 Identity in the Inf Soc'y 55.

[95] Zuboff, *supra* note 5 at 63-92.

[96] Tremblay, *supra* note 52 at 351.

deemed intent as a function of actual or deemed knowledge about the constituent parts of a platform's system (such as the five factors discussed in this paper).

iii. The Platform's Business Model

One of the reasons that technology companies have had so much difficulty recognizing, acknowledging, or meaningfully addressing emergent systemic harms arising from their platforms is because in many cases the harm can be traced back to the very business model that the company relies on to make money.[97] Perhaps the most well-known and prevalent example of this is the ad-based model that platforms such as Twitter, Facebook, Instagram, and Google rely on. Under this model, platform users enjoy what superficially appears to be the central services offered by each platform, such as reading and sharing articles, posting and commenting on photos, online search, engaging in public discussion, or sending private messages. Users ostensibly do not pay any fees for such services, but rather pay with their personal data and attention, both of which are harnessed in the service of third-party advertisers. According to Shoshanna Zuboff, it is these advertisers which are the true customers of such online platforms, whose core services are advertising, with the user-oriented services merely a supply chain means to serve the primary ends of data-based targeted advertising and associated behavioural prediction and influence.[98]

The important thing to remember is that this business model was legal when first introduced and for many years subsequent, nor did general public sentiment consider it outright wrong or unethical (albeit still cause for concern). While regulators in some jurisdictions have since imposed constraints on ad-based platform companies, such as ensuring users' informed consent,[99] this only regulates edges of the business model—and some would argue ineffectually· where regulation does not touch the heart of the model itself.[100] The exchange of users' personal data and exposure to targeted advertising for social media, short-term accommodation, ride-sharing, and other platform services remains a broadly legal and acceptable business proposition that continues alive and kicking to this day.

Despite this business model's legality and, until recently, wide public acceptability, the ad-based platform model has played a fundamental role in several emergent systemic harms that have since become established in mainstream awareness, thanks to the work of academics and researchers such as Zeynep Tufekci, Frank Pasquale, and Shoshanna Zuboff. For example, in her bluntly titled TED Talk, "We're building a dystopia just to make people click on ads", Tufekci explains how an ad-driven business model only incentivizes platforms to serve up to users the online content equivalent of increasingly unhealthy junk food, regardless of broader longterm consequences for public discourse or democracy.[101] Similarly, Zuboff demonstrates that by setting up a business model where all personal data becomes one endless supply chain to feed a demand for "prediction products" that will increase the effectiveness of targeted advertising, platform companies have in fact charted out a course whose logical end results in the loss of

[97] Zuboff, *supra* note 5 at 104.

[98] Zuboff *supra* note 5 at 10.

[99] See e.g. EC, *Commission Regulation (EC) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, [2018] OJ, L 119/1 [GDPR].

[100] This parallels Hughes' assertion that "'the regulatory model' of Canadian environmental law does not prohibit environmental harm; it merely regulates it: '[T]he entire approach is based upon setting up a hierarchy of degradation of (violence again) nature. We do not seek to protect nature from all harm, because of its inherent value. Instead, we seek to regulate how much harm is done, stopping short only when we might harm our own self-interests. As a result some damage is simply allowed … other harm is deemed to be unacceptable but in practice is condoned." McLeod-Kilmurray, "Ecofeminist Legal Critique", *supra* note 16 at 145.

[101] Zeynep Tufekci, "We're building a dystopia just to make people click on ads" (September 2017), online (video): TED <www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads>.

personal autonomy and citizens' individual and collective "right to the future tense".[102] Again, this business model has been legal, and was not generally considered inherently abhorrent at first blush—yet nonetheless led to several systemic harms that experts such as Tufekci, Pasquale, and Zuboff have shown may be existentially threatening to human autonomy and dignity, as well as to fundamental pillars of a free and democratic society itself.

iv. Time and Scale of the Platform's and Users' Activities

The abovementioned effects of YouTube's recommendation algorithms may not be so critical—albeit would remain significant and concerning—if such recommendations were largely ignored by users, or if YouTube itself were a bit player with a tiny user base. However, YouTube's impact resulted from two additional factors key to emergent systemic harm: time and scale. Even if all of the above-described factors were in play for a particular platform—business model with negative externalities; differentially perceived technological affordances; and algorithms optimized too well for the wrong things or with collateral damage—the platform itself may not give rise to systemic harms until reaching a particular threshold of critical mass, in the form of number of users and reach. One of the main reasons that the major online platforms have garnered so much attention—e.g., Facebook (including Instagram and WhatsApp), Google (including YouTube), Amazon, Airbnb, and Uber—compared to their smaller, more niche, or less popular peers (e.g., Vimeo, Twitch, Tumblr) is precisely due to their sheer size and reach,[103] thus influence and tangible impact on the fabric of society.

YouTube's recommendation engine is a core part of viewers' experience on the platform, and a core component of the company's business model. According to one journalist, "Company insiders tell me the algorithm is the single most important engine of YouTube's growth",[104] reportedly "driv[ing] 70 percent of total time on the platform."[105] Google's own researchers stated in a paper, "YouTube represents one of the largest scale and most sophisticated industrial recommendation systems in existence."[106] This level of influence—or *choice governance*—over such a large proportion of users (who then become political actors beyond the platform) "makes the platform's demonstrated tendency to jump from reliable news to outright conspiracy all the more worrisome."[107] While scale is not a surefire metric by which to assess the potential likelihood of systemic harm—just think of the devastation that certain 4chan or 8chan boards have wreaked, for instance, despite their relatively smaller user base—this only means that the

[102] Zuboff, *supra* note 5 at 196.
[103] According to 2018 Pew research, for instance, YouTube is a regular destination for 73% of all U.S. adults and 94% of U.S. young adults (age 18-24): Aaron Smith & Monica Anderson, "Social Media Use in 2018" (1 March 2018), online: Pew Research Center <www.pewinternet.org/2018/03/01/social-media-use-in-2018/>. In Canada, YouTube enjoyed over 560 million monthly visits as of January 2019, from 24 million Canadians (among a total 2 billion monthly logged in users globally), or nearly 70% of adults in Canada: J Clement, "Leading websites in Canada as of January 2019, by average monthly traffic (in million visits)" (11 September 2019), online: Statista <(https://www.statista.com/statistics/1047699/canada-websites-ranking-by-average-monthly-traffic/>.; Irene S Berkowitz, Charles H Davis & Hanako Smith, "Watchtime Canada: How YouTube Connects Creators and Consumers" (22 May 2019), online (pdf): *Ryerson University Faculty of Communication & Design: Audience Lab* <sites.google.com/view/watchtime-2019https://drive.google.com/file/d/1frDtTaR-y7qHQuECH-4M83zOhfBk-t2Z/view>; and Chris Powell, "Nearly 70% of Canadians Watch YouTube Monthly (Report)" (11 November 2015), online: *Marketing* <marketingmag.ca/media/nearly-70-of-canadians-watch-youtube-monthly-report-161240/>.
[104] Lewis, *supra* note 71.
[105] Fisher & Taub, "Brazil", *supra* note 70.
[106] Paul Covington, Jay Adams & Emre Sargin, "Deep Neural Networks for YouTube Recommendations" (Paper published in the Proceedings of the 10th ACM Conference on Recommender Systems, ACM, New York, 2016), online (pdf): <storage.googleapis.com/pub-tools-public-publication-data/pdf/45530.pdf>.
[107] O'Donovan et al, *supra* note 71.

absence of scale is not necessarily cause for reassurance, while the presence of scale should certainly be considered a red flag for potential emergent systemic harm.

Often, time is related to scale, in the sense that some of today's most problematic platforms, such as Google and Facebook, required time to build up to the impact that they currently have, in terms of user base, resources, technological capacity, and political influence. This may be why, for example, discussions around CDA 230 can be so fraught: the impact of the provision has changed over time, as the platforms it protects has changed over time. Despite these significant evolutions (or devolutions, depending on perspective), various interlocutors approach intermediary liability debates as if the platforms have always existed as they do today, or on the contrary, still exist today as they did when the legislation first came into force in 1996, when neither is the case. What this suggests is that regulators, lawmakers, stakeholders, and other decision-makers should monitor online platforms on an ongoing basis, and document how their operations and impacts change over time, and evaluate potential emergent impacts accordingly.

On the other hand, sometimes it is *lack* of time that results in emergent systemic harm, rather than the passage of time. The most likely instance of this is in cases where online platforms have seen a sudden explosion of popularity, and achieve scale in a short period of time, with resulting off-platform impacts that surrounding environments were not prepared to handle. For example, the navigation app Waze drew negative legal[108] and political[109] attention when it ended up diverting massive volumes of daily rush-hour traffic, seemingly overnight, off of major highways and through formerly quiet residential areas that were not equipped to handle the traffic, nor were given the time to make any necessary accommodations.[110]

v. Human Nature and Users' Individualist Instrumentalism

The final major constituent element of emergent systemic harms arising from digital platforms is, for lack of better terminology, the atomization, individualization, and instrumentalization in how users view their own activities and engage on digital platforms. For the most part, "people will use the platform as they see fit, but will not take on this same sense of ownership [as more dedicated and collective interest-minded users who take on volunteer content moderation and community policing roles within a platform]; for them the platform is instrumental, a way to distribute their videos or find their friends or discuss the news."[111] The public interest benefits that might flow from an online platform in a best-case scenario—such as democratizing media, facilitating equality-seeking social justice movements, or increasing diversity of representation across a wide range of fields—fall victim to, or are overshadowed by, a variant of the tragedy of the commons, or a failure to heed Kant's categorical imperative.[112]

[108] See e.g. Ela Levi-Weinrib, "Waze sued for directing drivers to quiet streets", *Globes* (1 December 2016), online: <en.globes.co.il/en/article-waze-sued-for-directing-drivers-to-quiet-streets-1001164377>.
[109] See e.g. Lisa W Foderaro, "Navigation Apps Are Turning Quiet Neighborhoods Into Traffic Nightmares", *New York Times* (24 December 2017), online: <www.nytimes.com/2017/12/24/nyregion/traffic-apps-gps-neighborhoods.html>; Henry Grabar, "Suburbs Finally Figured Out a Way to Get Rid of Pesky Drivers on Waze Shortcuts", *Slate* (16 June 2017), online: <slate.com/business/2017/06/suburbs-finally-figured-out-a-way-to-get-rid-of-pesky-drivers-on-waze-shortcuts.html>.
[110] With thanks to Rebekah Overdorf at KU Leuven for bringing this example to my attention.
[111] Gillespie, *Custodians*, *supra* note 10 at 93. Gillespie continues, "[A]s these platforms begin to stand in for the public itself, people no longer see them as a specific and shared project, to be joined and contributed to."
[112] "[A]ct only in accordance with that maxim through which you can at the same time will that it become a universal law." The Formula of the Universal Law of Nature" (7 July 2016), online: *Stanford Encyclopedia of Philosophy* <plato.stanford.edu/entries/kant-moral/#ForUniLawNat>.

For example, one systemic harm that has emerged from social media platforms such as Facebook and Twitter has been the proliferation and pervasiveness of disinformation and "fake news".[113] After accounting for each of the contributing factors of emergent systemic harm discussed above—i.e., each platform's ad-driven business model, technological affordances of easy sharing and reacting, algorithmic promotion of sensational content, and sheer scale of their user bases built up over time—at some point, the success of "fake news" must on some level also be attributed to readers' credulity, lack of media literacy, partisanship, and the human desire to share news and information that has prompted an emotional response. Sharing and reacting to content one encounters online are of course more than reasonable actions for private individuals to engage in, and in fact they are encouraged to do so by the platforms themselves. As well, it is neither wrong, *per se*, nor illegal to lack media literacy or to share and react to articles according to one's political views, as a private individual on a personal account. The law also does not require that every private individual act at all times in the interest of the public good, or evaluate every personal action for its potential systemic impact. However, the fact remains that a critical mass of users acting at scale, as atomized individuals isolated from their collective impact, may eventually give rise to an emergent systemic harm that none of the users themselves necessarily intended or would have agreed to if they had been presented with the option upfront.

To return to Waze as an example, it is precisely individuals acting in their own isolated interests—as they are allowed to by law, invited to by the app's design, and directed to by Waze's algorithm—that results in a net negative for all involved (given the research that after a certain scale is reached, traffic congestion merely spreads rather than is circumvented, and users obtain increasingly diminishing benefits).[114] One news article's title sums up this state of affairs as "The Perfect Selfishness of Mapping Apps".[115] Similarly, "gig economy" platforms such as Uber and Airbnb rely on individual drivers, passengers, hosts, and guests to benefit from using each platform to their own individual ends, regardless of broader systemic impacts that emerge as a result of each platform's popularity and knock-on repercussions. In the case of Airbnb, its popularity has resulted in exacerbating housing shortages and rental affordability in cities already struggling to address the issue,[116] while some studies have shown that Uber increased the number of vehicles on roads by pulling users off of public transit into private cars, as opposed to its advertised benefit of decreasing vehicles by encouraging pre-existing drivers to join carpools.[117]

[113] See e.g. Gabrielle Lim, "Disinformation Annotated Bibliography" (May2019), online (pdf): *Citizen Lab* <citizenlab.ca/wp-content/uploads/2019/05/Disinformation-Bibliography.pdf>; Fenwick McKelvey & Elizabeth Dubois, "Computational Propaganda in Canada: The Use of Political Bots" (2017), online: *Oxford Internet Institute* <comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Canada.pdf>; Soroush Vosoughi, Deb Roy & Sinan Aral, "The spread of true and false news online", *Science* (9 March 2018), online: <science.sciencemag.org/content/359/6380/1146>.

[114] Théophile Cabannes et al, "The Impact of GPS-Enabled Shortest Path Routing on Mobility: A Game Theoretic Approach" (Presented at *Transportation Research Board 97th Annual Meeting*, Washington DC, 7 January 2018).

[115] Alexis C Madrigal, "The Perfect Selfishness of Mapping Apps", *Atlantic* (15 March 2018), online: <www.theatlantic.com/technology/archive/2018/03/mapping-apps-and-the-price-of-anarchy/555551/>.

[116] See e.g. Tom Cardoso & Matt Lundy, "Airbnb likely removed 31,000 homes from Canada's rental market, study finds", *Globe and Mail* (21 June 2019), online: <www.theglobeandmail.com/canada/article-airbnb-likely-removed-31000-homes-from-canadas-rental-market-study/>; Max Fawcett, "Airbnb Versus Everyone", *Walrus* (23 July 2019), online: <thewalrus.ca/airbnb-versus-everyone/>; David Wachsmuth & Alexander Weisler, "Airbnb and the Rent Gap: Gentrification Through the Sharing Economy" (2018) 50:6 Env & Planning A: Economy and Space 1147.

[117] "TNCs tell a good-news story about how TNCs benefit urban America. They declare that their competition is the personal auto, not public transit. [...] TNCs have pushed back against the narrative that they promote automobility and unsustainably increase traffic congestion while also weakening public transportation. [...] These results clearly show that instead of 'replacing the personal auto,' TNCs in large cities are primarily supplanting more space-efficient modes such as bus, subway, biking and walking." Bruce Schaller, "The New Automobility: Lyft, Uber and the Future of American Cities" (25 July 2018), online (pdf): *Schaller Consulting* <www.schallerconsult.com/rideservices/automobility.pdf> at 15.

While it is true that we should "stop blaming technology for the failings of human beings"[118]—whether that is Facebook and Twitter for disinformation, Waze for traffic congestion and disrupting urban planning, YouTube for hate speech and conspiracy theories, or Google Search for perpetuating discriminatory stereotypes—there is a difference between holding platform companies responsible for human nature itself, and holding them responsible for *exploiting* human nature in the pursuit of commercial gain, regardless of the consequences.[119] It is not human nature alone, but human nature in conjunction with or as leveraged by all of the above-discussed constituent elements of online platforms—several of which platforms themselves may control—that can combine to create emergent systemic harm: a platform's business model, its technological affordances, its algorithms, and the scale of users' activities combined with either the passage or lack of time.

The components discussed above do not form necessarily an exhaustive list, but are a compilation of key factors that have played a role in known emergent systemic harms from online platforms and that are likely to play a role in future emergent systemic harms. Not all of these factors are necessarily required for systemic harm to emerge, nor does the presence of all of these factors necessarily mean that systemic harm will emerge for certain. Their presence, however, should put the platform company, users, stakeholders, lawmakers, and other relevant decision-makers on notice so that they remain alert to the potential of emergent systemic harm and can monitor and respond accordingly, so as to prevent some of the worst consequences that have already occurred to date as a result of insufficient recognition, acknowledgement, or awareness.

C. Examples of Platform-Facilitated Emergent Systemic Harm to a Marginalized Group

This section will bring together the above five components in two examples of "unintended consequences" that have resulted from online platforms and their users' activities, which fit the definition of emergent systemic harm: Google Search prompting auto-fill suggestions and providing search results based on harmful stereotypes; and YouTube's recommendation engine systematically leading viewers ever deeper into the recesses of right-wing extremism, while separately also unwittingly catering to paedophiles by recommending videos of children. Note that in these cases, the example is only considered *emergent* systemic harm before knowledge of the potential for harm was brought to the attention of the companies; after that point, the harm remains systemic, but is no longer emergent, as the platforms then knowingly continue to facilitate the harm. Additionally, Airbnb and ridesharing apps such as Uber and Lyft occupy a grey zone due to the variation in municipal laws and prohibitions. The remainder of this section will discuss each of these examples in turn.

In the case of Google Search, the constituent components were an ad-driven business model requiring user clicks to maximize revenues; the technological affordance of auto-filled search suggestions imparting a certain truth value to the suggestions; the search algorithm pushing purportedly popular user destinations to the top of search results, thus signifying a certain truth value to them in the eyes of online searchers; the sheer scale of Google Search making it a *de facto* arbiter of truth and information in the world; and the human nature of users actively searching for and/or often clicking on derogatory results concerning marginalized groups, such as results sexually objectifying black women and girls, or photos

[118] Mike Masnick, "We Should Probably Stop Blaming Technology for the Failings of Human Beings", *Techdirt* (3 July 2019), online: <www.techdirt.com/articles/20190616/02114442409/we-should-probably-stop-blaming-technology-failings-human-beings.shtml>.
[119] "What we are witnessing is the computational exploitation of a natural human desire: to look 'behind the curtain,' to dig deeper into something that engages us. As we click and click, we are carried along by the exciting sensation of uncovering more secrets and deeper truths. YouTube leads viewers down a rabbit hole of extremism, while Google racks up the ad sales." Tufekci, "Great Radicalizer", *supra* note 6.

criminalizing black teenagers.[120] None of the first four elements, on their own, are inherently wrong or unreasonable. The fifth element may be wrong, unethical, or condemnable, but in the context of private, personal searches, it is also not necessarily illegal. However, all five factors coalesced to produce the emergent systemic harms to racialized and especially black communities that Noble details in *Algorithms of Oppression*, without Google intending any such consequences before they were brought to the company's attention.

The case of YouTube also involves the same five elements, each on its own "innocuous", reasonable, legitimate, or at least legal, but combined with reality and the other elements becomes twisted into instruments of emergent systemic harm to various vulnerable populations (whether the targets of right-wing extremism, or children whose videos were watched and shared by paedophiles). The five constituent elements were: an ad-driven business model requiring maximizing viewers and viewing time to maximize profit; the technological affordances of auto-playing suggested videos upon the current one finishing, and of easily following recommended videos down extremist rabbit holes through presenting recommended videos alongside the currently watched one; the recommendation algorithms tying together extremist or otherwise harmful videos, that would have been isolated, into a network of content essentially giftwrapped for captive viewers; the sheer scale of YouTube granting the platform and its algorithms and users sociopolitical impact beyond the platform itself; and what Tufekci describes as the social media version of humans' vestigial need for excess salt, fat, and sugar, or human attention tending to favour the extreme, the sensational, the outrageous, and the partisan, independent of other aspects such as the presence or absence truth, accuracy, or impact on the quality of public discourse and democratic participation. Again, YouTube may not have intended any of the emergent systemic harms, to marginalized and vulnerable populations, that arose from its platform, but questions regarding the company's responsibility and potential liability arise after the point at which YouTube was informed of the demonstrated or likely consequences and did not act to mitigate or stop them.[121]

Airbnb's destructive impact on housing and rental markets and Uber's negative impact on public transit and the environment occupy a grey zone of emergent systemic harm. In municipalities where there were no pre-existing laws or regulations prohibiting short-term private accommodation rentals or unlicensed ridesharing, one could argue that Airbnb and Uber were doing nothing wrong and breaking no laws in any of the factors contributing to emergent systemic harm, including their business model; platform affordances inviting guests, hosts, drivers, and passengers to engage and use the service; or algorithms matching listings and drivers to users, while the hosts and drivers were themselves doing nothing wrong or illegal in offering unlicensed or untaxed short-term accommodations and driving services. (This is putting aside labour issues associated with Uber and Lyft drivers and their formal status as independent contractors; [122] such issues would also not be considered *emergent* systemic harms due to their being the result of informed and intentional decisions by the platform companies in the context of pre-existing legal frameworks.) Where such laws exist, however, then Airbnb's and/or Uber's users would be engaging in illegal activity, facilitated by the platforms, and the systemic harm they caused would no longer be considered emergent. However, for the purposes of determining liability for platform-facilitated systemic harm in a jurisdiction *without* such laws, it may be appropriate to take into account whether Airbnb or Uber knew or ought to have known that there were laws prohibiting their and their users' activities in other jurisdictions, and thus were put on notice to consider the rationales behind those laws and reflect on

[120] Noble, *supra* note 7 at 64-109.

[121] Mark Bergen, "YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant", *Bloomberg* (2 April 2019), online: <www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant>.

[122] See e.g. Kristine Owram, "Uber drivers are employees, not contractors, Canadian lawsuit argues", *Financial Post* (23 January 2017), online: <business.financialpost.com/transportation/uber-drivers-are-employees-not-contractors-canadian-lawsuit-argues>.

whether their platforms and users would foreseeably cause the systemic harm meant to be addressed by the prohibitive laws in other jurisdictions.

Not all emergent systemic harms to a marginalized group may include a marginalized group that is expressly protected under human rights law, but this is something that could be argued in a case at court. For example, socioeconomic status is not currently a recognized protected ground under section 15 of the *Charter*, meaning Airbnb's impact on housing may not meet the definition of harm to a legally protected marginalized group, unless listed under human rights legislation. However, social sciences evidence might show that low-income and would-be tenants are a marginalized class that is negatively impacted by the systemic harm of Airbnb's upheaval of the rental market, in a way that meets the definition of adverse effect.[123] In this case, a low-income tenants' association, for example, may have grounds to bring action against Airbnb for platform-facilitated systemic harm to a marginalized community.

Having established and defined some key constituent elements of emergent systemic harm arising from online platforms, as well as provided examples of what may constitute a platform-related emergent systemic harm, Part II of this paper will demonstrate how platform-facilitated emergent systemic harm to marginalized communities can and should be recognized as a legally actionable harm in Canadian law.

Part II. Establishing Emergent Systemic Harm to Marginalized Communities as a Legally Actionable Harm

Examining legal doctrines and principles that form the core of Canadian human rights and environmental law, respectively, provides support for recognizing online platform-facilitated emergent systemic harm to marginalized groups as a legally actionable harm in Canada. Canadian human rights and environmental law have each, in their own way, established legal recognition of and remedy for systemic harm. In the case of Canadian human rights law, federal and provincial human rights codes as well as jurisprudence under the *Canadian Charter of Rights and Freedoms* ("the *Charter*") have recognized that systemic discrimination is a scourge throughout society that results in devastating consequences for victimized groups.[124] Environmental law has developed certain legal principles meant to prevent significant harm to ecological systems, including potential future harm. While other areas of law also recognize systemic harm—for instance, class action lawsuits have recognized systemic harm in the context of consumer protection law, privacy law, and product liability; and competition law aims to protect economic systems in specific industries—this paper will focus on applying the law, legal principles, and legal scholarship from human rights and environmental law. In part this is due to time and scope constraints; however, three reasons underly relying on human rights and environmental law in particular.

First, the focus of this paper is on emergent systemic harms to marginalized groups in particular; it thus makes sense to turn to a major established area of law that speaks directly to protecting marginalized and vulnerable groups in society, particularly on the basis of equality and non-discrimination, one of the key concerns at the heart of public interest consternation with online platform companies. Second, there are several parallels between environmental issues and the online platforms issues examined in this paper that suggest environmental law may productively lend itself to the analysis in this paper. For instance, both

[123] See e.g. *Sparks v Dartmouth/Halifax County Regional Housing Authority*, SCA No 02681, 330 APR 91, 119 NSR (2d) 91, 30 RPR (2d) 146 (NSCA).

[124] See e.g. Ontario Human Rights Commission, *Under suspicion: Research and consultation report on racial profiling in Ontario* (April 2017), online (pdf): <www.ohrc.on.ca/sites/default/files/Under%20suspicion_research%20and%20consultation%20report%20on%20racial%20profiling%20in%20Ontario_2017.pdf>; and National Inquiry into Missing and Murdered Indigenous Women and Girls, *Reclaiming Power and Place: The Final Report of the National Inquiry into Missing and Murdered Indigenous Women and Girls* (June 2019).

online platforms and the Internet they form a large part of (in practice for the average user, if not in technical fact) have been compared to intertwined, ecological systems, where events that occur in one part of the system ripple out to affect another part of the system, eventually reaching beyond the system itself (i.e. affect people and other types of systems, such as democratic elections). Third, environmental law incorporates an explicit element of pre-emptive regulation as a form of remedy for potential harms and risks even in the face of uncertainty; the increasing calls for pre-emptive regulation of online platforms today strongly warrants examining contexts where Canadian law has already implemented such measures.

In addition to the above reasons, areas of law such as consumer protection and competition have traditionally focused narrowly on economic harms and impacts, such as the price of consumer services. However, this narrow focus has come under increasing scrutiny with it becoming increasingly clear that poor competition among the Internet platform companies affects far more than consumer prices, while consumer price has become divorced from consumer and more importantly overall human welfare,[125] as seen in the formal price of social media services being zero, while the informal, unquantifiable price is something much greater. Both human rights law and environmental law share the common attribute of providing legal recognition and remedies for harms that are not easily measurable or quantifiable, but go to the core of fundamental concerns such as human dignity and the sustainable well-being of our planet.

A. Human Rights Laws Protect Marginalized Communities from Systemic Harm

The concept of systemic harm is uncontroversial in Canadian human rights law (not to be confused with debate regarding how to establish systemic harm or whether a given activity contributes to systemic harm). The few judicial and tribunal decisions that mention the specific phrase "systemic harm" do not explicitly give it a legal definition, but rather refer to it in passing as a presumed known plain language term, particularly in the context of equality rights and racial or other forms of discrimination.[126]

In this paper, systemic harm to marginalized groups is equated with systemic discrimination, a concept long established in federal and provincial human rights law, and which may also be known by the terms adverse effect discrimination, adverse impact discrimination, or indirect discrimination (in contrast to direct, overt, or intentional discrimination). Protection against systemic discrimination is enshrined in statutory human rights codes and has been reinforced by decisions from the Supreme Court of Canada. The following provisions in the *Canadian Human Rights Act* sets out characteristics that are protected from discrimination:

> Prohibited grounds of discrimination
>
> 3 (1) For all purposes of this Act, the prohibited grounds of discrimination are race, national or ethnic origin, colour, religion, age, sex, sexual orientation, gender identity or expression, marital status, family status, genetic characteristics, disability and conviction for an offence for which a pardon has been granted or in respect of which a record suspension has been ordered.
>
> (2) Where the ground of discrimination is pregnancy or child-birth, the discrimination shall be deemed to be on the ground of sex.
>
> (3) Where the ground of discrimination is refusal of a request to undergo a genetic test or to disclose, or authorize the disclosure of, the results of a genetic test, the discrimination shall be deemed to be on the ground of genetic characteristics.[127]

[125] See generally Lina Khan, "Amazon's Antitrust Paradox" (2017) 126 Yale LJ 710.

[126] See e.g. *Fontaine v Canada (Attorney General)*, 2013 ONSC 684, at paras 18 and 28; *Canada (Attorney General) v The Canadian Broadcasting Corp*, 2016 ONSC 4938*; Campbell v Jones*, 2001 NSSC 139 at page 5.

[127] *Canadian Human Rights Act*, RSC 1985, c H-6, s 3.

Similarly, the *Charter* sets out the following grounds of protection: "Every individual is equal before and under the law and has the right to the equal protection and equal benefit of the law without discrimination and, in particular, without discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability."[128]

Discrimination and systemic discrimination are further defined as follows—while the quoted excerpt speaks of the employment context in particular, the description of systemic discrimination and its feedback-loop effects for those discriminated against can also apply to contexts beyond employment, such as access to services or political participation:

> Although Judge Abella chose not to offer a precise definition of systemic discrimination, the essentials may be gleaned from the following comments, found at p. 2 of the Abella Report:
>
> > "Discrimination ... means practices or attitudes that have, whether by design or impact, the effect of limiting an individual's or a group's right to the opportunities generally available because of attributed rather than actual characteristics ….
> >
> > It is not a question of whether this discrimination is motivated by an intentional desire to obstruct someone's potential, or whether it is the accidental by-product of innocently motivated practices or systems. If the barrier is affecting certain groups in a disproportionately negative way, it is a signal that the practices that lead to this adverse impact may be discriminatory.
> >
> > This is why it is important to look at the results of a system …."
>
> In other words, systemic discrimination in an employment context is discrimination that results from the simple operation of established procedures of recruitment, hiring and promotion, none of which is necessarily designed to promote discrimination. The discrimination is then reinforced by the very exclusion of the disadvantaged group because the exclusion fosters the belief, both within and outside the group, that the exclusion is the result of "natural" forces, for example, that women "just can't do the job."[129]

With respect to identifying when providing legal remedy for a systemic issue is appropriate, versus individual harms, Canadian jurisprudence on the law of public interest standing in the context of constitutional criminal cases is instructive, specifically *Chaoulli v Quebec (AG),* 2005 SCC 35 ("*Chaoulli*"), and *Canada (Attorney General) v. Downtown Eastside Sex Workers United Against Violence Society*, 2012 SCC 45 ("*DTES SWUAVS*"). In *DTES SWUAVS,* the Supreme Court of Canada determined that the plaintiffs involved needed to be able to address *systemic* issues in the criminal law related to sex work offences, that they had "raised issues of public importance *that transcend their immediate interests*", and that challenging the legislative regime as a whole "may prevent a multiplicity of individual challenges in the context of criminal prosecutions."[130] The Court went on to state, "The

[128] *Charter*, *supra* note 26, s 15(1).

[129] *CN v Canada (Canadian Human Rights Commission)*, [1987] 1 SCR 1114 at pages 1138-39; See similar definitions of systemic discrimination in Ontario ("Systemic discrimination results when a neutral rule has a disproportionate negative impact on individuals because they belong to a certain group or possess certain characteristics.", *Lawrence v IBEW, Local 120* (1986), [1986] OLRB Rep 1241, at paras 24, 31) and PEI ("Systemic discrimination arises from the existence of a particular policy which creates the discriminatory effect. The effect is usually obvious (e.g. height and weight restrictions). Systemic discrimination arises out of long-standing stereotypes and value assumptions which create the discriminatory effect ... systemic discrimination is quite often the result of unintentional behaviour. There is no desire to exclude certain people or classes of people but, as the result of stereotypes, mind-sets and attitudes which have been acquired over a long period of time, the effect is discriminatory." *Ayangma v Prince Edward Island (2001)*, 2001 PESCAD 22, [2001] PEIJ No 105, 108 ACWS (3d) 979, at para 26).

[130] *Canada (Attorney General) v Downtown Eastside Sex Workers United Against Violence Society*, 2012 SCC 45 [*DTES SWUAVS*] at para 73.

presence of the individual respondent, as well as the Society, will ensure that there is both an individual and collective dimension to the litigation."[131] Moreover, the BC Court of Appeal, whose decision the Supreme Court affirmed, drew directly on Binnie and LeBel JJ.'s dissent in *Chaoulli c. Québec (Procureur general)*, "characterizing the *Charter* challenge in that case as a 'systemic' challenge, which differs in scope from an individual's challenge addressing a discrete issue."[132] This attentiveness to the collective dimension of an issue and to stakes that transcend narrow individual interests is what is required in a more complete analysis of the harms that arise from online platforms and associated users' activities.

The BC Court of Appeal (BCCA) in *DTES SWUAVS* also elaborated on what it means for an issue to be "systemic":

> The term "systemic" is something of a chameleon: it is used where an entire legislative scheme is challenged and, particularly in human rights cases, is used to describe situations in which disproportionately adverse consequences accrue to persons from legislative provisions that do not, on their face, target those persons adversely. […] The differences between a systemic challenge and an individual direct challenge, particularly in cases alleging discrimination, was the subject of comment by this Court in *British Columbia v. Crockford*, 2006 BCCA 360 (CanLII), 271 D.L.R. (4th) 445 at para. 49.
>
> > "[49] A complaint of systemic discrimination is distinct from an individual claim of discrimination. Establishing systemic discrimination depends on showing that practices, attitudes, policies or procedures impact disproportionately on certain statutorily protected groups: see *Radek* at para. 513. A claim that there has been discrimination against an individual requires that an action alleged to be discriminatory be proven to have occurred and to have constituted discrimination contrary to the Code. The types of evidence required for each kind of claim are not necessarily the same. Whereas a systemic claim will require proof of patterns, showing trends of discrimination against a group, an individual claim will require proof of an instance or instances of discriminatory conduct."[133]

Platform-facilitated emergent systemic harms to marginalized groups arise precisely because a platform's business "practices, attitudes, policies or procedures impact disproportionately on certain statutorily protected groups". Whether it is Facebook's "neutral" "Real Name" policy hurting *mostly* drag queens, sex workers, or abuse survivors; Google's "neutral" search engine hurting *mostly* black teenagers and black women and girls; YouTube's "neutral" recommendation algorithms consistently leading viewers overwhelmingly into right-wing extremist content that spills over into hate speech and hate crimes in real life, which target *mostly* immigrants, women, and LGBTQ+ individuals; or Airbnb's "neutral" system creating negative externalities that disproportionately impact *mostly* low-income housing-seekers—digital platforms facilitate a number of harms that would likely fall under the definition of systemic discrimination in Canadian human rights law.

Furthermore, the BCCA majority in *DTES SWUAVS* granted public interest standing to the plaintiffs because "the essence of the complaint was that the law impermissibly renders individuals vulnerable while they go about otherwise lawful activities and exacerbates their vulnerability."[134] This characterization is also the essence of many charges against online platforms, when it comes to "exacerbating the vulnerability" of users who already belong to vulnerable populations, as well as non-users who are impacted by repercussions of activity on an online platform that reverberate beyond the platform itself. Coordinated campaigns of online abuse and harassment that leverage platform features in

[131] *Ibid*.
[132] *Ibid* at para 15.
[133] *Downtown Eastside Sex Workers United Against Violence Society v Canada (Attorney General)*, 2010 BCCA 439, at paras 58-60.
[134] *DTES SWUAVS*, *supra* note 130 at para 16.

unexpected ways to target women, members of the LGBTQ+ community, and Black, Indigenous, and other racialized persons are the foremost among many examples of this dynamic. The crucial aspect of an analysis focusing on systemic harm, rather than individual harms, is that "[s]uch a structural effect need not be the product of individual actors—although they play a role—but rather can operate systemically."[135]

Another critical component of systemic discrimination is that it does not have to be intentional. This legal prioritization of impact over intent is rooted in Canada's federal and provincial statutory human rights codes, which do not require complainants to demonstrate that the defendant intentionally engaged in discrimination, only that discrimination occurred. See, for example, the British Columbia Human Rights Code: "Discrimination in contravention of this Code does not require an intention to contravene this Code."[136] The emphasis of effect over intent or motive has also been established under Ontario human rights law: "An intention to discriminate is not required."[137] The Supreme Court of Canada has similarly recognized that "[d]iscriminatory intent on behalf of an employer is not required to demonstrate *prima facie* discrimination: *Bombardier*, at para 40."[138] Thus, emergent systemic harm to marginalized groups facilitated by platform companies, even in the absence of intent on the part of platforms, can still constitute legally recognizable harm warranting liability and remedy.

Moreover, the Supreme Court of Canada has established that the protected ground does not necessarily have to be the sole or primary cause of discrimination, but merely a contributing factor:

> [F]or a particular decision or action to be considered discriminatory, the prohibited ground need only have contributed to it […]
>
> A close relationship is not required in a discrimination case under the *Charter*, however. To hold otherwise would be to disregard the fact that, since there may be many different reasons for a defendant's acts, proof of such a relationship could impose too heavy a burden on the plaintiff. Some of those reasons may, of course, provide a justification for the defendant's acts, but the burden is on the defendant to prove this. It is therefore neither appropriate nor accurate to use the expression "causal connection" in the discrimination context. […]
>
> [T]he plaintiff has the burden of showing that there is a *connection* between a prohibited ground of discrimination and the distinction, exclusion or preference of which he or she complains or, in other words, that the ground in question was a *factor* in the distinction, exclusion or preference.[139]

That a protected ground only has to be a "contributing factor" rather than the sole or central factor to justify access to legal remedy is important due to the difficulty in definitively proving discrimination given what is nearly always a state of information and evidentiary asymmetry. Sheppard and Chabot discuss how "one of the most difficult obstacles for plaintiffs in establishing *prima facie* discrimination is proving a connection between individual exclusion or disadvantage and one or more group-based ground(s) of discrimination."[140] This difficulty persists despite the fact that "the essence of discrimination is mistreatment of an individual based on his or her affiliation with a particular group or groups".[141]

[135] Michelle Y Williams, "African Nova Scotian Restorative Justice: A Change Has Gotta Come" (2013) 36:2 Dal LJ 419, citing *Report of the Royal Commission on the Donald Marshall, Jr, Prosecution* (1989).
[136]*Human Rights Code*, RSBC 1996, c 210, s 2.
[137] *MacDonald v London Health Sciences Centre*, 2019 HRTO 1134, at para 96, citing *Ontario (Human Rights Commission) v Simpson-Sears Ltd*, [1985] 2 SCR 536 at para 18 and *ADGA Group Consultants Inc v Lane*, 91 OR (3d) 649, 295 DLR (4th) 425, 240 OAC 333 (ON Div Ct), at para 153.
[138] *Stewart v Elk Valley Coal Corp*, 2017 SCC 30 [*Elk Valley*] at para 24.
[139] *Quebec (Commission des droits de la personne et des droits de la jeunesse) v Bombardier Inc (Bombardier Aerospace Training Center)*, 2015 SCC 39, at paras 48-52.
[140] Colleen Sheppard & Mary Louise Chabot, "Obstacles to Crossing the Discrimination Threshold: Connecting Individual Exclusion to Group-Based Inequalities" (2018) 96:1 Can Bar Rev 1 at 14.
[141] *Ibid*.

Sheppard and Chabot point to informational and evidentiary asymmetry as one main reason for this challenge, where defendants are usually in sole possession of the knowledge and evidence behind the facts of their actions leading to discriminatory impact.[142] This dynamic has led some human rights tribunals, such as in the specific case of *Clennon v. Toronto East General Hospital,*[143] to be "willing to infer *prima facie* discrimination in the absence of a non-discriminatory explanation",[144] because in many cases, "only the respondent knew the true reasons" for their actions.[145]

Information asymmetry is notoriously severe and rampant in the imbalanced relationships between online platforms and their users (let alone non-users).[146] Similar reasoning should thus apply in finding systemic harm to marginalized groups arising from the activities or business models of online platforms. Further, Sheppard and Chabot draw a parallel between discrimination cases and medical liability cases because in both situations, "the defendant tends to be the best person to explain what happened since they have more knowledge of and direct involvement in the process leading to the result."[147] This lends greater support to establishing platform-facilitated systemic harm to protected groups as a legally recognized harm in tort, as some principles central to medical liability have already been applied to the online platform context as a potential way to hold platform companies accountable, such as through the concept of a fiduciary duty.[148]

To remedy this difficulty of information asymmetry and ensure that Canadian human rights and equality law fulfills its purpose, Sheppard and Chabot argue that courts and tribunals must give due weight to "broader evidence of social context, including systemic realities of racism, stereotyping and unconscious bias" playing a role in defendants' actions leading to discriminatory impacts. They point to "compelling examples [of this broader evidence] being relied upon to assist judges and adjudicators in understanding the likelihood of discrimination."[149] This broader approach to evidence that recognizes the deep influence of social context should also inform analyses of systemic harms that arise from online platforms, particularly given the numerous studies that have demonstrated how such social context impacts the experiences of different groups of users characterized by protected grounds.[150]

[142] *Ibid* at 16.

[143] *Clennon v Toronto East General Hospital*, 2009 HRTO 1242.

[144] Sheppard & Chabot, *supra* note 140 at 17.

[145] *Ibid* at 16.

[146] Zuboff, *supra* note 5 at 192 (where such asymmetry is part of what Zuboff terms the "unauthorized privatization of the division of learning in society"; Pavlovíc, "Contracting Out", *supra* note 14 at 399 ("Consumer contracts, however, embody an inherent power imbalance, which necessitates greater intervention by the state.").

[147] Sheppard & Chabot, *supra* note 140 at 19.

[148] See e.g. Ian Kerr, "Personal Relationships in the Year 2000: Me and My ISP" in Law Commission of Canada, ed, *Personal Relationships of Dependence and Interdependence in Law* (Vancouver: UBC Press, 2002) 78; and Jack M Balkin, "Information Fiduciaries and the First Amendment" (2016) 49:4 UC Davis L Rev 1185.

[149] Sheppard & Chabot, *supra* note 140 at 27. The authors add the following qualification: "While it is always important not to base one's conclusions exclusively on group-based generalizations—a phenomenon that ironically is often at the root of discrimination itself—when used carefully, an understanding of societal inequalities and patterns of exclusion allows adjudicators to make certain inferences of discrimination. And of course, any factual inferences may still be refuted by defendants, who in most instances have greater knowledge of the underlying bases of their decisions to exclude or deny equal treatment." *Ibid* at 27 (footnotes omitted).

[150] See e.g. Anne Helen Petersen, "The cost of reporting while female", *Columbia Journalism Review* (2018), online: <https://www.cjr.org/special_report/reporting-female-harassment-journalism.php>; "Women journalists feel the brunt of online harassment" (April 2019), online: *Mozilla* <https://internethealthreport.org/2019/women-journalists-feel-the-brunt-of-online-harassment/>; Megan Farokhmanesh, "Google's LGBTQ Employees Are Furious About YouTube's Policy Disasters", *Verge* (7 June 2019), online: <https://www.theverge.com/2019/6/7/18656540/googles-youtube-lgbtq-employees-harassment-policies-pride-month>.

B. Environmental Law Principles Focus on Preventing Systemic Harm

Addressing systemic harm in the online platform environment can take inspiration from environmental law principles. After all, the Internet, online platforms, ubiquitous connectivity, and mobile devices have given rise to their own digital ecosystems, an information environment, and various public commons of sorts (albeit largely under the control and governance of private entities). That many also think of online environments in terms similar to the natural environment is apparent in the relevant literature, research, policy, and media coverage of online platform issues.[151]

For example, Safiya Noble draws a direct connection between the two contexts, in an interview where she rejected the idea that users should be solely responsible for "filtering out false, manipulative and toxic" content online: "That's like saying the public should adapt to their water being poisoned. If the information environment is poisoned, that's not on the public at large to solve."[152] Similarly, a UNESCO handbook on journalism and disinformation states that "strong ethical journalism is needed as an alternative, and antidote, to the contamination of the information environment and the spill-over effect of tarnishing of news more broadly";[153] and speaks of a disinformation "polluting the information environment".[154] The rise of smart-city initiatives has also led van der Graaf and Ballon to characterize affected urban spaces as "complex platform-based ecosystem[s] encompassing private and public organizations and people/citizens".[155] While many of these examples refer to disinformation—which still disproportionately impacts marginalized communities[156]—Sarah Jeong identified early on that harassment

[151] See e.g. Timur Ablyazov & Viktoriya Rapgof, "Digital platforms as the basis of a new ecological system of socioeconomic development" (Paper published in *IOP Conference Series: Materials Science and Engineering*, 2019), online (pdf): <iopscience.iop.org/article/10.1088/1757-899X/497/1/012002/pdf>; Danny Bradbury, "Fake news: Mozilla joins the fight to stop it polluting the web" (22 August 2017), online: *Naked Security* <nakedsecurity.sophos.com/2017/08/22/fake-news-mozilla-joins-the-fight-to-stop-it-polluting-the-web/>; Antonio Cartelli, "From Smart Cities to Smart Environment: Hints and Suggestions for an Ecology of the Internet" (2012) 3:4 Intl J of Digital Literacy and Digital Competence 7. In addition, the Internet Society has defined the "Internet Ecosystem" as a term to indicate that "the rapid and continued development and adoption of Internet technologies can be attributed to the involvement of a broad range of actors; open, transparent, and collaborative processes; and the use of products and infrastructure with dispersed ownership and control"—again pointing to the idea of collective action and systemic interactions comprising online environments, rather than siloed individual actors. "The Internet Ecosystem" (February 2014), online (pdf): *Internet Society* <www.internetsociety.org/wp-content/uploads/2017/09/factsheet_ecosystem.pdf>.

[152] Emily Cavalcanti, "The search for ethics" (17 October 2018), online: *USC Annenberg School for Communication and Journalism* <annenberg.usc.edu/news/feature/search-ethics>..

[153] UNESCO, Cherilyn Ireton & Julie Posetti, eds, *Journalism, 'Fake News' & Disinformation: Handbook for Journalism Education and Training* (2018), online (pdf): <en.unesco.org/sites/default/files/journalism_fake_news_disinformation_print_friendly_0.pdf> at 9.

[154] *Ibid* at 18. See also mentions of "polluting", "polluters", and "pollution" of information, media, and political environments throughout Yochai Benkler, Robert Faris & Hal Roberts, *Network Propaganda* (Oxford: Oxford University Press, 2018).

[155] Shenia van der Graf & Pieter Ballon, "Navigating Platform Urbanism" (2019) 142 Tech'l Forecasting & Soc Change 364 at 365. See also Anne Faber, Florian Matthes & Felix Michel, eds, "Digital Mobility Platforms and Ecosystems: State of the Art Report" (July 2016), online (pdf): *Technical University of Munich* <mediatum.ub.tum.de/doc/1324021/file.pdf>.

[156] See e.g. ob Faris et al, "Partisanship, Propaganda, & Disinformation: Online Media & the 2016 US Presidential Election" (August 2017), online (pdf): *Berkman Klein Center for Internet & Society* <cyber.harvard.edu/sites/cyber.harvard.edu/files/2017-08_electionReport.pdf>; Amanda Morris, "Right-wing WhatsApp users in Brazil are more effective at spreading disinformation", *Phys* (29 August 2019), online: <phys.org/news/2019-08-right-wing-whatsapp-users-brazil-effective.html>; Yochai Benkler et al, "Study: Breitbart-led right-wing media ecosystem altered broader media agenda", *Columbia Journalism Review* (3 March 2017), online: <www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>; Dan Kennedy, "A Major New Study

itself is a form of Internet pollution that prevents those targeted and harassed from being able to fully engage in and enjoy the benefits of online environments, in her aptly named book, *The Internet of Garbage*.[157]

The discursive recognition of the Internet and online platforms as constituting ecosystems, ecological networks, and shared environments akin to the natural environment, combined with environmental law's focus on systemic analysis and system-level harms, suggests that laws purporting to remedy emergent systemic harm from online platforms may adopt certain key principles from environmental law. According to DeMarco, the Supreme Court of Canada has upheld several "fundamental environmental values" throughout its environmental law decisions.[158] Addressing systemic harms from platform algorithms should consider incorporating two of these values especially: the "polluter pays" principle and the precautionary principle.[159]

i. Polluter/Beneficiary Pays

The "polluter pays" principle, as articulated by the Supreme Court of Canada in *Imperial Oil*, "assigns polluters the responsibility for remedying contamination for which they are responsible and imposes on them the direct and immediate costs of pollution."[160] Polluters are also "asked to pay more attention to the need to protect ecosystems in the course of their economic activities."[161] The Court stated that this principle "has become firmly entrenched in environmental law in Canada [and] is found in almost all federal and provincial environmental law in Canada".[162] According to DeMarco, this universal adoption of "polluter pays" across Canada means the concept can be applied in many ways across litigatory, legislative, and regulatory contexts, including in allocating tort liability.[163] He argues that not only does the principle assist in determining liability, it also broadens the scope of what a defendant may be found liable *for*, such as "environmental losses that are not [strictly] reflected in commercial or market terms".[164] This understanding of non-quantifiable losses will likely also be instrumental in establishing platform-facilitated emergent systemic harms (whether to marginalized groups or not), as they often cannot be appropriately quantified or reduced to monetary damages.

The "polluter pays" principle has already been making its way into the legal and policy discourse surrounding online platforms. For example, in defending a decision against Google establishing the "right to be forgotten" at the European Court of Justice, the United Kingdom's then-Information Commissioner

Shows That Political Polarization Is Mainly A Right-Wing Phenomenon", *WGBH News* (15 March 2017), online: < https://www.wgbh.org/news/2017/03/15/politics-government/major-new-study-shows-political-polarization-mainly-right-wing>; "The Human Consequences of Computational Propaganda Eight Case Studies from the 2018 U.S. Midterm Elections" (2019), online: *Institute for the Future* <www.iftf.org/disinfoeffects>; Craig Silverman & Jane Lytvynenko, "Vulnerable Groups Could Be Targeted And Silenced Online Ahead Of 2020 Election, Researchers Warn", *BuzzFeed* (7 May 2019), online: <www.buzzfeednews.com/article/craigsilverman/extremists-disproportionally-target-and-silence-latinos>; Nicola Pardy, "How The Fake News Industry Weaponizes Women", *Refinery29* (3 May 2018), online: <www.refinery29.com/en-us/2018/05/197259/misinformation-fake-news-misogyny-sexism-women-discrimination>.

[157] Sarah Jeong, *The Internet of Garbage* (Vox Media, 2018), at pages 68-68, 71-72.

[158] These are "environmental rights, the polluter pays principle, the precautionary principle, intergenerational equity, sustainability, and public trust". Jerry V DeMarco, "The Supreme Court of Canada's Recognition of Fundamental Environmental Values: What Could be Next in Canadian Environmental Law?" (2007) 17:3 J Envtl L & Prac 159.

[159] See McLeod-Kilmurray, "Ecofeminist Legal Critique", *supra* note 16, for an analysis of critical application of courts' application of these two principles in the context of GMOs, patent law, and environmental law.

[160] *Imperial Oil Ltd v Quebec (Minister of the Environment)*, 2003 SCC 58 at para 24 [*Imperial Oil*].

[161] *Ibid*.

[162] *Ibid* at para 23.

[163] DeMarco, *supra* note 158 at 182.

[164] *Ibid* at 183, citing *British Columbia v Canadian Forest Products Ltd*, 2004 SCC 38.

stated, "The polluter pays, the polluter should clear up. Google is a massive commercial organisation making millions and millions out of processing people's personal information. They're going to have to do some tidying up."[165] The Carnegie UK Trust has proposed applying a statutory duty of care to online platforms, stating that such a duty "returns the cost of harms to those responsible for them, an application of the micro-economically efficient 'polluter pays' principle."[166] During a parliamentary debate in 2018, Lord Stevenson asked, "If it is right to operate a 'polluter pays' principle, whereby the costs of pollution prevention and control measures are met by the polluter, why is that principle not equally valid in the social media companies?"[167]

A variant of the "polluter pays" principle is "beneficiary pays", under which some suggest any beneficiary of the pollution be deemed a "polluter", to address situations where the actual polluter in a situation was not the same entity that benefited.[168] "Beneficiary pays" has been defined to involve the following principle: "a person who benefited from the activity resulting in the contamination should share liability for its cleanup with other responsible persons", or "the notion that those who benefit economically from an activity that creates environmental hazards should pay for its cleanup."[169] This principle, which has appeared in some Canadian court cases,[170] may be more useful in context of emergent systemic harms from online platforms, where the platform itself may not have intentionally or directly "polluted" their online environment for marginalized communities, but nonetheless benefits from the actions and dynamics the platform facilitates which leads to the systemic harm. For example, Google did not deliberately set out to establish pornography as the top-ranking search results for black girls and other racialized women and girls; however, the search engine clearly benefited from both the searches themselves and the algorithms that produced such results. The search engine had adverse effects on racialized women based on discriminatory attitudes that objectified, exoticized, and sexualized them as a group based on their gender and race, which polluted the "public" information environment governed by search results, even if Google did not intend for that to occur. Under this analysis, racialized women may have grounds for a cause of action against Google for facilitating, through its search engine platform, emergent systemic harms to them based on characteristics protected under Canadian human rights and equality laws.

ii. Precautionary Principle

The precautionary principle is known as follows:

> In order to protect the environment, the precautionary approach shall be widely applied by States according to their capability. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation.[171]

[165] Kelly Fiveash, "ICO: It's up to Google the 'POLLUTER' to tidy up 'right to be forgotten' search links", *Register* (24 July 2014), online: <www.theregister.co.uk/2014/07/24/ico_chief_says_google_needs_to_tidy_up_right_to_be_forgotten_requests_as_search_engines_meet_brussels_officials/>.

[166] "Response to The Online Harms White Paper" (June 2019), online (pdf): *Carnegie UK Trust* <d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/07/04163920/Online-Harm-White-paper-.pdf> at 4.

[167] UK, HL Deb (12 November 2018), vol 793 (Lord Stevenson of Balmacara).

[168] Nickie Vlavianos, "Creating Liability Regimes for the Clean-up of Environmental Damage: The Literature" (1999) 9 J Envtl L & Prac 145 at 154.

[169] *Ibid* at 155.

[170] See e.g. *Kawartha Lakes (City) v Ontario (Director, Ministry of the Environment)*, [2009] OERTD No 59, 48 CELR (3d) 95, at para 63 and Appendix C: Excerpts of Canadian Council of Ministers of the Environment "Contaminated Site Liability Report: Recommended Principles for a Consistent Approach Across Canada" (1993), Principle 4; *Gehring v Chevron Canada Ltd*, 2007 BCCA 557 at para 14.

[171] UN, *Rio Declaration on Environment and Development*, 1992, 31 ILM 874, Principle 15.

The precautionary principle is a cornerstone of international environmental law standards and environmental legislation in Canada.[172] This principle has been recognized by the Supreme Court as well as in "numerous lower court and tribunal proceedings" throughout Canada.[173] Demarco has suggested that this near-universal recognition and incorporation of the principle may be used to enforce a duty on decision-makers to act, or refrain from acting, as adherence to the principle would demand.[174] Even more so than polluter-pays, calls for accountability from online platform companies and regulatory proposals have often included reference to the precautionary principle, either by name or in spirit.[175] However, some argue that it would be inappropriate to apply the precautionary principle in the context of technology, drawing on arguments rooted in the social, political, and literal capital of "innovation".[176] While such arguments may have made sense for online platforms and technologies in their infancy, and when current familiar dynamics today were unprecedented, today we are better equipped to identify potential contributing factors to emergent systemic harm to marginalized communities from online platforms, and are more aware of the demonstrated consequences, to an extent that may warrant earlier intervention compared to before. Where there remains true uncertainty with a brand new platform, the company introducing the risk for its own benefit should bear the burden of demonstrating or ensuring that it will not result in systemic harm to marginalized communities. The burden of demonstrating harm should not fall upon the communities themselves, either in the form of expending resources for advocacy and litigation, or in the form of disproportionately suffering systemic harm that in fact comes to pass.

Part III. "Unintended Consequences" and the Platform Foreseeability Gap

Emergent systemic harms from platform algorithms are frequently labelled, framed, or downplayed as "unintended consequences",[177] with the implicit notion that such consequences were thus also unforeseen by the associated platform company (and thus could not have prevented them). Indeed, the label of "unintended consequences" as applied to platform algorithm-facilitated harms was what guided the selection of the particular examples discussed throughout and what formed the basis of the analysis in the first half of this paper. Before advancing to applying a legal test that turns on foreseeability, however, the "unintended" nature of such harms warrants interrogation. Part III of this paper breaks down and

[172] DeMarco, *supra* note 158 at 186.

[173] *Ibid*.

[174] *Ibid* at 186-190.

[175] See e.g. Sarah E Light, "Precautionary Federalism and the Sharing Economy" (2017) 66 Emory LJ 333; Lorna Woods & William Perrin, "Online harm reduction – a statutory duty of care and regulator" (April 2019), online: *Carnegie UK Trust* <repository.essex.ac.uk/25261/1/Online-harm-reduction-a-statutory-duty-of-care-and-regulator.pdf>; David Souter, "Inside the Information Society: Permissionless innovation and the precautionary principle" (2 April 2018), online: *Association for Progressive Communications* <www.apc.org/en/blog/inside-information-society-permissionless-innovation-and-precautionary-principle>.

[176] See e.g. Joshua New & Daniel Castro, "How Policymakers Can Foster Algorithmic Accountability" (21 May 2018), online (pdf): *Center for Data Innovation* <www2.datainnovation.org/2018-algorithmic-accountability.pdf>; Adam Thierer, "Avoiding a Precautionary Principle for the Internet", *Forbes* (11 March 2012), online: <www.forbes.com/sites/adamthierer/2012/03/11/avoiding-a-precautionary-principle-for-the-internet/#7a8b577e7ff2>.

[177] See e.g., "And on it goes, with each company and technology platform producing its own graveyards full of unintended consequences. Facebook disseminates journalism widely but ends up promoting vacuous and sometimes politically pernicious clickbait. Google works to make information (including the content of books) freely available to all but in the process dismantles the infrastructure that was constructed to make it possible for people to write for a living. Twitter gives a megaphone to everyone who opens an account but ends up amplifying the voice of a demagogue-charlatan above everyone else, helping to propel him all the way to the White House." Damon Linker, "The genius and stupidity of Silicon Valley", *The Week* (20 October 2017), online: <https://theweek.com/articles/731764/genius-stupidity-silicon-valley>.

distinguishes between "unintended", "unforeseen" in various permutations (e.g., unforeseen but only at an earlier point in time, unforeseen but only by the platform and outright predicted by others), and "unforeseeable" (including unforeseeable by *whom*). These distinctions result in what may be termed the "platform foreseeability gap", as well as destabilization of the concept of emergence as applied to platform-facilitated systemic harms. This shifting nature of emergence is then carried forward into the discussion of applying reasonable foreseeability in this context, under the negligence test in tort law.

A. The Importance of Being Unintended

The rhetoric of "unintended consequences" has persisted throughout media, academic, and political discourse, research, and opinion concerning, for example:

- Facebook contributing to the decline of journalism and the news industry;[178]
- Facebook enabling employment and housing discrimination through targeted advertising;[179]
- Facebook's role enabling foreign interference in elections and impaired electoral integrity[180]
- the negative impact of Uber and Lyft on urban traffic and public transit ridership;[181]

[178] See e.g., "On top of this, platforms like Facebook will often fail to gauge the consequences of their actions … In response, Facebook modified their algorithms to focus on friends posts rather than news. An unintended consequence was that many legitimate news outlets saw their revenue plummet." R Cetina Presuel & JM Martínez Sierra, "Algorithms and the News: Social Media Platforms as News Publishers and Distributors" (2019) 18:2 Revista de Comunicación 261 (inline citations omitted); "The worry for publishers is that such an approach will have the unintended consequence of hurting high-quality content because a lot of legitimate news articles, while they may get read, tend not to get shared or commented on." Lucia Moses, "'We're losing hope': Facebook tells publishers big change is coming to News Feed" (11 January 2018), online: *Digiday* <https://digiday.com /media/losing-hope-facebook-tells-publishers-big-change-coming-news-feed/>; "Facebook doesn't *care* about news — it cares about getting people to use Facebook, and its enormous effects on the news business has been an unintended consequence." Jillian D'Onfro, "Facebook is telling the world it's not a media company, but it might be too late" (29 August 2016), online: *Yahoo Finance* <https://finance.yahoo.com/news/facebook-telling-world-not-media-185848782.html> (emphasis in original).

[179] See e.g., "Several of the union's Facebook's ads, purchased in November, feature videos and photos of female members and women of color; the ads are not targeted by gender, but by location and to those interested in construction. Nevertheless, Maher said, the audience reached by the ads is about two-thirds men. Paradoxically, because advertisers can no longer target by age or gender, they have little recourse to remedy these disproportions. "It's an unintended consequence," Mislove said. "You can't say steer it toward men instead. Facebook gives you no way to say, 'I want this to be balanced.'" Ava Kofman & Ariana Tobin, "Facebook Ads Can Still Discriminate Against Women and Older Workers, Despite a Civil Rights Settlement", *Pro Publica* (13 December 2019), online: <https://www.propublica.org/article/facebook-ads-can-still-discriminate-against-women-and-older-workers-despite-a-civil-rights-settlement>; "Facebook said it will also study the potential for unintended bias in its algorithms, and will meet with the plaintiff groups every six months for the next three years to discuss implementation of the changes." Josh Eidelson, Sarah Frier & Bloomberg, "Facebook to Block Discriminatory Ads in 'Historic' Legal Accord", *Fortune* (19 March 2019), online: <https://fortune.com/2019/03/19/facebook-discriminatory-ads/>;

[180] Bret Stephens, "Facebook's Unintended Consequence", *New York Times* (3 May 2019), online: <https://www.nytimes.com/2019/05/03/opinion/facebook-free-speech.html>.

[181] See e.g., "The services have changed the way people get around the city and, for some, have opened up an opportunity to earn extra income as a driver. But they may also have an unintended consequence: more traffic congestion." Mary Wisniewski & Jennifer Smith Richards, "Uber and Lyft said ride-sharing would cut traffic congestion and supplement public transit. But has it worked?" *Chicago Tribune* (7 June 2019), online: <www.chicagotribune.com/news/breaking/ct-biz-ride-share-congestion-loop-20190520-story.html>; and "Lyft's introduction will eat into transit ridership — an often overlooked, but unintended consequence of ride-hailing services." Bern Grush, "The future of transit is ride sharing and driverless cars", *The Star* (20 November 2017), online: <www.thestar.com/opinion/contributors/2017/11/20/the-future-of-transit-is-ride-sharing-and-driverless-cars.html>.

- the impact of Waze on traffic congestion in small neighbourhoods;[182]
- the anti-competitive implications of Amazon's market stratagems;[183]
- YouTube's role in galvanizing networks of white supremacist extremism and conspiracy theorists;[184]
- YouTube's recommendation engine efficiently catering to paedophiles;[185]
- Airbnb enabling discrimination against racialized guests and hosts;[186]
- Airbnb's destructive effects on local residential rent vacancies and affordable housing;[187]

[182] See e.g., "The new design is meant to be safer for pedestrians, better for transit and to discourage speeding by drivers. An unintended consequence has been an increase in drivers using mapping software that sends them down the narrow streets that traverse the neighborhood's steep hillsides, said resident Katherine Murphy. She's angry at the city for presuming everyone would play nice and only use the main intersection. 'Do they just have their heads in the sand that people are using technology?' she said." Elizabeth Weise, "Waze and other traffic dodging apps prompt cities to game the algorithms", *USA Today* (6 March 2017), online: <www.usatoday.com/story/tech/news /2017/03/06/mapping-software-routing-waze-google-traffic-calming-algorithmsi/98588980/>; and The cut-through disputes "are an unintended consequence of this great technology that is supposed to help people avoid sitting in traffic," said Paul Silberman, a traffic engineer with Sabra, Wang & Associates in Columbia, Md. Steve Hendrix, "Traffic-weary homeowners and Waze are at war, again. Guess who's winning?" *Washington Post* (5 June 2016), online: < https://www.washingtonpost.com/local/traffic-weary-homeowners-and-waze-are-at-war-again-guess-whos-winning/2016/06/05/c466df46-299d-11e6-b989-4e5479715b54_story.html>.

[183] "I think what we have right now is the unintended consequence. The giants are destroying competition in one area after another. It's the job of law to enforce the rules to stop them." Cat Zakrzewski, "Elizabeth Warren: 'It's not even hard' to break up Amazon and other tech giants", *Los Angeles Times* (11 March 2019), online: <www.latimes.com/business/la-fi-tn-warren-amazon-tech-breakup-20190311-story.html>

[184] See e.g., this quote from a YouTube's executive regarding its recommendation algorithm's "rabbit hole effect": "Yeah, so I've heard this before, and I think that there are some myths that go into that description that I think it would be useful for me to debunk. The first is this notion that it's somehow in our interests for the recommendations to shift people in this direction because it boosts watch time or what have you. I can say categorically that's not the way that our recommendation systems are designed. […] I can also say that it's not in our business interest to promote any of this sort of content." Kevin Roose, "YouTube's Product Chief on Online Radicalization and Algorithmic Rabbit Holes", *New York Times* (29 March 2019), online: <www.nytimes.com/2019/03/29/ technology/youtube-online-extremism.html>; and "Bergen's story [about YouTube] is, in a way, a mirror of the *New York Times*' November story on how Facebook first ignored, then sought to minimize warning signs about the platform's unintended consequences. Both pieces illustrate the ugly fashion in which our social networks have developed: Phase one is an all-out war to gain user attention and build an advertising business; phase two is a belated effort to clean up the many problems that come with global scale faster than new ones can arise. Casey Newton, "How extremism came to thrive on YouTube", *Verge* (3 April 2019), online: <https://www.theverge.com/ interface/2019/4/3/18293293/youtube-extremism-criticism-bloomberg>.

[185] See e.g., "YouTube never set out to serve users with sexual interests in children — but in the end, Mr. Kaiser said, its automated system managed to keep them watching with recommendations that he called 'disturbingly on point.'" Fisher & Taub, "Open Gate", *supra* note 69.

[186] See e.g., "We find that non-black hosts are able to charge approximately 12% more than black hosts, holding location, rental characteristics, and quality constant. Moreover, black hosts receive a larger price penalty for having a poor location score relative to non-black hosts. These differences highlight the risk of discrimination in online marketplaces, suggesting an important unintended consequence of a seemingly-routine mechanism for building trust." Benjamin Edelman & Michael Luca, "Digital Discrimination: The Case of Airbnb" (10 January 2014), Harvard Business School Working Paper 14-054, online (pdf): <https://hbs.edu/faculty/Publication%20Files /Airbnb_92dd6086-6e46-4eaf-9cea-60fe5ba3c596.pdf>.

[187] See, e.g., "Airbnb, launched as a way for property owners (or renters) to make some money and travelers to save some, has for the last several years been under fire for an unintended consequence: Pricing out renters and threatening affordable housing." Lark Turner, "On Affordable Housing, is Airbnb the Problem or the Solution?" (17 June 2015) Urban Turf: <https://dc.urbanturf.com/articles/blog/on_affordable_housing_is_airbnb_the_problem_ or_the_solution/10012>.; and "The company has shifted the burden of rising prices in crowded downtown areas from travelers to residents—pushing down prices for hotel rooms, while raising rents for city dwellers. Was that

- Google Search providing racist, sexist, and otherwise discriminatory search results;[188]
- the role of social media platforms in the proliferation of disinformation and the impaired state of public discourse, particularly with respect to political and sociocultural issues;[189]
- hate speech and related abuse on Facebook, Twitter, and other social media platforms;[190] and

Airbnb's intent? Almost certainly not. But that is the outcome, anyway, and it is a meaningful—even, yes, disruptive—one. […] Like just about every story these days about revolutionary tech platforms, Airbnb is a story both of democratized access to commerce and the unintended consequences of those democratizing efforts, even when they succeed on their own terms." Derek Thompson, "Airbnb and the Unintended Consequences of 'Disruption,'" *Atlantic* (17 February 2018), online: <https://www.theatlantic.com/business/archive/2018/02/airbnb-hotels-disruption/553556/>.

[188] See e.g., "Google Maps gave racist, degrading results not because it was compromised, but because the internet itself is racist and degrading. That revelation comes from a Google statement posted yesterday evening. 'Certain offensive search terms were triggering unexpected maps results, typically because people had used the offensive term in online discussions of the place,' wrote Jen Fitzpatrick, VP of Engineering & Product Management. 'This surfaced inappropriate results that users likely weren't looking for.'" Brian Barrett, "Google Maps Is Racist Because the Internet Is Racist", *Wired* (23 May 2015), online: <https://www.wired.com/2015/05/google-maps-racist/>.

[189] See e.g., "It's clear that the spread of misinformation was an unintended consequence of the deployment of algorithms to maximize engagement. Social media platforms—very understandably—followed the money trail. But a lack of foresight doesn't absolve these companies of culpability, Noble says." Katherine J Wu, "Radical ideas spread through social media. Are the algorithms to blame?" *PBS Nova* (28 March 2019), online: <https://www.pbs.org/wgbh/nova/article/radical-ideas-social-media-algorithms/>; "As web companies strive to tailor their services (including news and search results) to our personal tastes, there's a dangerous unintended consequence: We get trapped in a 'filter bubble' and don't get exposed to information that could challenge or broaden our worldview." Eli Pariser, "Beware online 'filter bubbles'" (March 2011), online (video): TED <https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles>; "Some even argue that the social networks are easy to flood with disinformation by design - an unintended consequence of their eagerness to cater to advertisers by categorising the interests of their users." "Facebook apologises for racist ad", *Daily Mercury* (4 November 2018), online: <https://www.dailymercury.com.au/news/facebook-apologises-for-category-promoting-white-s/3566980/>; "One unintended consequence of commercially effective social platform design is that it undermines the architecture of knowledge and healthy discourse." Emily Bell, "Facebook can no longer be 'I didn't do it' boy of global media" (11 November 2016), online: *Columbia Journalism Review* <https://www.cjr.org/tow_center/facebook_zuckerberg_trump_election.php>; and "This kind of machine learning can produce unintended effects, with worrying consequences for public discourse. In the past 18 months, many have wondered whether Facebook's algorithms have deepened political divisions and facilitated the spread of misinformation and propaganda." Alex Abdo, "Facebook is shaping public discourse. We need to understand how", *Guardian* (15 September 2018), online: <https://www.theguardian.com/commentisfree/2018/sep/15/facebook-twitter-social-media-public-discourse>.

[190] See e.g., "Name a problem that you have encountered on Twitter, or an unintended consequence of Twitter that you worry about, and Dorsey will cop to it. He is talking, but one reason that he is talking to let you know that he is listening, too. He has been listening for a long time." Newton, "Twitter's CEO", *supra* note 40; "It is perhaps unsurprising then that an unintended consequence of digital technology has been the extent to which some individuals and groups have used the freedom to participate online to engage in hateful or discriminatory communicative practices in these loosely regulated spaces..." Emily Harmer & Karen Lumsden, "Online Othering: An Introduction" in K Lumsden & E Harmer (eds), *Online Othering: Exploring Violence and Discrimination on the Web* (Basingstoke: Palgrave Macmillan, 2019); and "Facebook recently announced that group content will receive more prominent placement in the News Feed and that groups overall will be a core product focus to help reverse an unprecedented decline in active users in the US and Canada. The unintended consequence is that the more than a billion users active in groups are being placed on a collision course with the hackers, trolls, and other bad actors who will follow Facebook's lead and make groups even more of a focus for their activities." Craig Silverman, Jane Lytvynenko & Lam Thuy Vo, "How Facebook Groups Are Being Exploited To Spread Misinformation, Plan Harassment, And Radicalize People", *BuzzFeed News* (19 March 2018), online: <https://www.buzzfeednews.com/article/craigsilverman/how-facebook-groups-are-being-exploited-to-spread>.

- the role of social media platforms in the potential downfall of democracy and, in one instance, humanity itself.[191]

Implicit in this rhetoric is the notion that because these harmful consequences were ostensibly unintended and unintentional, they were not only unfore*seen* (by the platform companies), but unforesee*able* (by anyone). Closer examination of platform-facilitated emergent systemic harms shows that neither of these conclusions was quite true in all cases.

i. Unintended but Foreseen by Platform

First, numerous leaks and revelations from investigative reporting have demonstrated that online platforms companies, in some cases, knew more than they had let on with respect to the role of their respective technologies and businesses in bringing about particular adverse consequences. For example, Facebook appears to have known of Cambridge Analytica scraping user data from its platform at a much earlier point in the relevant timeline, such as when management ignored, then waved off the concerns of, an employee who "requested clarification on Facebook's policies for political consultancies that were scraping data to match Facebook profiles to the lists of voters that campaigns use, known as voter files" and requested an investigation into "what Cambridge specifically is actually doing".[192] Facebook was also found to have "detected elements of the Russian information operation [to interfere in the United States 2016 presidential election through coordinated disinformation campaigns on Facebook] in June 2016", yet dismissed the theory as "crazy" in 2017 before its own internal investigation confirmed otherwise.[193] Similarly, YouTube steadfastly ignored warnings and concerns from among its own ranks, about the platform's role in promoting disinformation, abusive content, and right-wing extremism,[194] as well as in facilitating and tolerating the popularity of inappropriate and traumatic content involving children.[195]

[191] "An indication for the extent of the problem can be seen by Facebook's own admission that *social media can have the unintended consequence of corroding democracy*." Antonis Matakos & Aristides Gionis, "Tell me Something My Friends do not Know: Diversity Maximization in Social Networks" (Paper delivered at the IEEE International Conference on Data Mining, 17-20 November 2018), online: *IEEE* <https://ieeexplore.ieee.org/ document /8594857>. (emphasis in original); "Here's what an AI apocalypse looks like in broad brushstrokes: 1. We carefully program AI with a well-defined goal for the betterment of the human race, give it power, and set it loose. 2. It develops unexpected emergent behaviours, and begins to have impacts unforeseen by its creators. 3. The well-defined goal is achieved. Unfortunately, the unintended consequence is the downfall of humanity. Let's compare this to current circumstances." Robert Smith, "Click here for the AI apocalypse (brought to you by Facebook)", *Guardian* (23 November 2016), online: <https://www.theguardian.com/commentisfree/2016/nov/23/ai-apocalypse-facebook-algorithms>.

[192] Julia Carrie Wong, "Document reveals how Facebook downplayed early Cambridge Analytica concerns", *Guardian* (23 August 2019), online: <www.theguardian.com/technology/2019/aug/23/cambridge-analytica-facebook-response-internal-document>.

[193] Adam Entous, Elizabeth Dwoskin & Craig Timberg, "Obama tried to give Zuckerberg a wake-up call over fake news on Facebook", *Washington Post* (24 September 2017), online: <https://www.washingtonpost.com/business /economy/obama-tried-to-give-zuckerberg-a-wake-up-call-over-fake-news-on-facebook/2017/09/24/15d19b12-ddac-4ad5-ac6e-ef909e1c1284_story.html>.

[194] "Wojcicki and her deputies know this. In recent years, scores of people inside YouTube and Google, its owner, raised concerns about the mass of false, incendiary and toxic content that the world's largest video site surfaced and spread. One employee wanted to flag troubling videos, which fell just short of the hate speech rules, and stop recommending them to viewers. Another wanted to track these videos in a spreadsheet to chart their popularity. A third, fretful of the spread of 'alt-right' video bloggers, created an internal vertical that showed just how popular they were. Each time they got the same basic response: Don't rock the boat." Bergen, *supra* note 121.

[195] "The company was facing an ongoing advertiser boycott, but the real catalyst was an explosion of media coverage over disturbing videos aimed at children. The worst was "Toy Freaks," a channel where a father posted videos with his two daughters, sometimes showing them vomiting or in extreme pain. YouTube removed Toy Freaks, and quickly distanced itself from it. But the channel hadn't been in the shadows. With over eight million

Additional cases that did not involve emergent systemic harms, but have also been characterized as "unintended consequences" arising from online platforms and their business decisions, also include some element of prior knowledge where the companies could have taken or refrained from action to prevent or mitigate the forecasted harmful consequences, and did not. For example, Uber was warned about the questionable safety of its self-driving cars shortly before one such car hit and killed a cyclist in Arizona;196 and Uber in Singapore "rented out faulty cars to their drivers, despite knowing that there had been a recall because they are a fire-risk", with one car bursting into flames shortly after the driver had dropped off a passenger.[197] Facebook "was repeatedly warned by its own employees as well as outsiders about a dangerous loophole that eventually led to the massive data breach in September 2018. Despite this, the loophole remained open for nine months after it was first raised,"[198] with Facebook protecting its own employees but not users from the known vulnerability.[199] The resulting harms from these egregious actions or omissions would fall outside the definition of "emergent" established in this paper, given the presence of known malpractice within a node in the system. However, they are worth mentioning to the extent they have contributed to or exacerbated the harms of allegedly "unintended" consequences. [200] Such examples complicate the notion of a given systemic harm being "emergent" in the way that Tremblay warns, and bring into question the nature of platform companies' intent.

ii. Unforeseen by Platform but Foreseen by Experts and Vulnerable Users

Second, even in cases where the platform companies had no internal inkling that their respective platforms' algorithms, business models, or affordances may lead to systemic harms, this did not mean that *no one* predicted potential outcomes. Such "unintended consequences", then, might be considered to have

subscribers, it had been reportedly among the top 100 most watched on the site. These types of disturbing videos were an 'open secret' inside the company, which justified their existence often with arguments about free speech, said one former staffer." *Ibid*.

[196] See e.g., "Uber told self-drive cars unsafe days before accident", *BBC News* (13 December 2018), online: <https://www.bbc.com/news/technology-46552604>; Jennings Brown, "Uber Employee Warned Self-Driving Cars 'Are Routinely in Accidents' Days Before Fatal Crash: Report", *Gizmodo* (11 December 2018), online: <https://gizmodo.com/uber-employee-warned-self-driving-cars-are-routinely-in-1831019048>.

[197] Sian Bradley, "Uber knowingly rented fire-risk cars to drivers in Singapore", *Wired* (4 August 2017), online: <https://www.wired.co.uk/article/uber-rent-car-singapore-fire>; Douglas MacMillan & Newley Purnell, "Smoke, Then Fire: Uber Knowingly Leased Unsafe Cars to Drivers", *Wall Street Journal* (3 August 2017), online: <https://www.wsj.com/articles/smoke-then-fire-uber-knowingly-leased-unsafe-cars-to-drivers-1501786430>.

[198] Laurence Dodds, "Facebook was repeatedly warned of security flaw that led to biggest data breach in its history", *Telegraph* (9 February 2020), online: <https://www.telegraph.co.uk/technology/2020/02/09/facebook-repeatedly-warned-security-flaw-led-biggest-data-breach/>.

[199] Katie Paul, "Facebook failed to warn users of known risks before 2018 breach: court filing", *Reuters* (15 August 2019), online: <https://www.reuters.com/article/us-facebook-privacy-lawsuit/facebook-failed-to-warn-users-of-known-risks-before-2018-breach-court-filing-idUSKCN1V600N>.

[200] For instance, Facebook deliberately propagating false video metrics (inflated by 150 to 900 percent), which potentially accelerated the "unintended consequence" of a downward-spiralling traditional journalism industry. Chris Welch, "Facebook may have knowingly inflated its video metrics for over a year", *Verge* (17 October 2018), online: <https://www.theverge.com/2018/10/17/17989712/facebook-inaccurate-video-metrics-inflation-lawsuit>; Alexis C Madrigal & Robinson Meyer, "How Facebook's Chaotic Push Into Video Cost Hundreds of Journalists Their Jobs", *Atlantic* (18 October 2018), online: <https://www.theatlantic.com/technology/archive/2018/10 /facebook-driven-video-push-may-have-cost-483-journalists-their-jobs/573403/>.; and Facebook not mitigating but itself weaponizing disinformation against its critics in a sustained offensive campaign: "Facebook employed a Republican opposition-research firm to discredit activist protesters, in part by linking them to the liberal financier George Soros. It also tapped its business relationships, lobbying a Jewish civil rights group to cast some criticism of the company as anti-Semitic"—fuelling pre-existing right-wing conspiracy theories against marginalized groups including racial justice activists. Sheera Frenkel et al, "Delay, Deny and Deflect: How Facebook's Leaders Fought Through Crisis", *New York Times* (14 November 2018), online: <https://www.nytimes.com/2018/11/14/ technology/facebook-data-russia-election-racism.html>.

been (at best, on the most generous reading) unforeseen by the platforms, but were by no means *unforeseeable*. The consequences were indeed foreseen, just not by those to whom the platform companies were listening. For example, technosociologist Zeynep Tufekci raised warnings about "the first Facebook-fueled ethnic cleansing campaign" in Myanmar as early as 2013,[201] five years before Facebook acknowledged the problem and took initial steps to address it.[202] Danielle Citron set out in 2007 considerations central to algorithmic decision-making today,[203] and Ian Kerr proposed potentially applying fiduciary obligations to online service providers as early as 2002.[204] Not only academics but industry experts as well have played Cassandra, as suggested by this one tweet in which the frustration is palpable:

> Amazon was lying about its video surveillance doorbell keeping people safe? Who knew? I mean... who could have predicted that an internet-enabled smart device would be a severe security risk? LIKE PERHAPS THE ENTIRE SECURITY INDUSTRY!!![205]

[201] Zeynep Tufekci, "I'd wondered in 2013 if Myanmar/Burma would be the first Facebook-fueled ethnic cleansing campaign. Here we are." (25 October 2017 at 10:05), online: *Twitter* <twitter.com/zeynep/status/923234138789990401>; Zeynep Tufekci, "I'm not a Myanmar/Burma expert. If I was worried as early as 2013, Facebook should have had a giant emergency team. I would like more info." (25 October 2017 at 10:37), online: *Twitter* <twitter.com/46eynep/status/923242101185433602>; and "zeynep tufekci tweeted 'This one gets me the maddest. Facebook had no excuse being so negligent about Myanmar. Here's me tweeting about it IN 2013. PEOPLE HAVE BEGGED FACEBOOK FOR YEARS TO BE PRO-ACTIVE IN BURMA/MYANMAR. Now he's hiring 'dozens'. This is a historic wrong.'" Adam Taylor, "Adam Taylor: Zuckerberg barely talked about Facebook's biggest global problem", *Salt Lake Tribune* (13 April 2018), online: <https://www.sltrib.com/opinion/commentary/2018/04/13/adam-taylor-zuckerberg-barely-talked-about-facebooks-biggest-global-problem/>.

[202] "It took until August 2018 – a year after 25,000 Rohingya were killed by Myanmar's army and allied Buddhist militias and 700,000 Rohingya were forced to flee the country – for Facebook to ban Tatmadaw leaders from its platform. Facing growing public pressure, the company also commissioned and published an independent Human Rights Impact Assessment on the role its services were playing in the country and committed to hiring 100 native Burmese speakers as content moderators." Julia Carrie Wong, "'Overreacting to failure': Facebook's new Myanmar strategy baffles local activists", *Guardian* (7 February 2019), online: <https://www.theguardian.com/technology/2019/feb/07/facebook-myanmar-genocide-violence-hate-speech>.

[203] "The twenty-first century's automated decision-making systems bring radical change to the administrative state that last century's procedural structures cannot manage. In the past, computer systems helped humans apply rules to individual cases. Now, automated systems have become the primary decision makers. These systems often take human decision making out of the process of terminating individuals' Medicaid, food stamp, and other welfare benefits. [...] Automation generates unforeseen problems for the adjudication of important individual rights. Some systems adjudicate in secret, while others lack recordkeeping audit trails, making review of the law and facts supporting a system's decisions impossible. […] The opacity of automated systems shields them from scrutiny. Citizens cannot see or debate these new rules. In turn, the transparency, accuracy, and political accountability of administrative rulemaking are lost. Code writers lack the properly delegated authority and policy expertise that might ameliorate such unintentional policymaking. They also usurp agency expertise when they inadvertently distort established policy." Danielle Keats Citron, "Technological Due Process" (2008) 85:6 Wash ULR 1249 at 1252-55.

[204] Ian Kerr, "Personal Relationships in the Year 2000: Me and My ISP" in Law Commission of Canada, ed, *Personal Relationships of Dependence and Interdependence in Law* (Vancouver: UBC Press, 2002) 78. See also Ian Kerr, "Good to see the information fiduciary concept finally gaining traction. Heads were shaking when I wrote about this concept for the Law Commission of Canada study of personal relationships of dependence and interdependence in 2000 https://static1.squarespace.com/static/56b8dbd62eeb817f29aa3265/t/576766d8c534a524caa2a747/1466394329423/PersonalRelationshipsintheYear2000-MeandMyISP.pdf" (10 April 2018 at 5:00), online: *Twitter* <https://twitter.com/ianrkerr/status/983812088157028354>.

[205] Brian McNett, "Amazon was lying about its video surveillance doorbell keeping people safe? Who knew? I mean... who could have predicted that an internet-enabled smart device would be a severe security risk? LIKE PERHAPS THE ENTIRE SECURITY INDUSTRY!!!" (10 December 2019), online: *Twitter* <https://twitter.com/b_mcnett/status/1204488820885835776>.

Beyond—though also overlapping with—academia and experts external to platform companies, Black women, particularly Black feminists, led and bore the brunt of being the early vanguard against the "meme warfare, fake news, foreign and domestic trolls" more widely endemic across online platforms such as Twitter today,[206] yet their concerns were ignored or dismissed by online platforms and traditional media alike.[207] More broadly, the technology sector's poor track record with diversity and equity issues among its employees and management[208] has been directly tied to such platforms' persistent inability to foresee what was and is, to others, the utterly predictable: "'The company needs to focus on listening to employees from marginalized backgrounds, many of whom could have warned it about many of the major disasters it's faced, many of whom tried to and weren't listened to,' said Irene Knapp, a Google employee of more than four years."[209] Similar points have been made with respect to now-mainstream concerns regarding widespread digital surveillance—conducted via digital platforms or otherwise—but which used to pass under the general public radar when mainly Black, Indigenous, and other racialized communities, as well as LGBTQ+ communities and those living with socioeconomic disadvantage, were its primary

[206] "Out of all online communities, Black Twitter and other black feminist spaces have been the most successful in repelling troll infiltration, memetic manipulation and calling out fake news wherever the source, be it from misguided Americans, raging hackers obsessed with Russia or other nations hostile to the US. An impressive feat given their adversary list, and one the community gets absolutely no credit for. [...] The years of living in this online abusive atmosphere has a price. Various sources describe this unpaid labor, and having to be being forever vigilant, as 'exhausting' and 'mentally draining.' On top of the fatigue, the harassment has only gotten worse now that the Republican party has embraced 4chan behaviors (like dogpiling), their language (the homosexual slur in the tweet embedded here) and memes like the joy of 'liberal tears.'" Fruzsina Eordogh, "Black Feminists Are USA's Best Defense Against Meme Warfare, Fake News, Foreign And Domestic Trolls", *Forbes* (9 March 2018), online: <https://www.forbes.com/sites/fruzsinaeordogh/2018/03/09/black-feminists-are-usas-best-defense-against-meme-warfare-fake-news-foreign-and-domestic-trolls/#8bd4be45de56>.

[207] See e.g., Flavia Dzodan, "This is why the work of Black women like @so_treu or @sassycrass was summarily dismissed as 'not research' or 'not scholarship' and instead, treated like 'internet debris' rather than quoted and respected. 5 years later journalists are still contending with their findings" (15 March 2019 at 3:51), online: *Twitter* <https://twitter.com/redlightvoices/status/1106462809871867904>; and Sydette Harry, "So when do we actually start talking about and paying BLACK WOMEN ESPECIALLY who predicted this and got laughed off" (15 November 2018 at 11:15), online: *Twitter* <https://twitter.com/Blackamazon/status/1063103216039018496> in reference to Color of Change, ".@facebook 's response to us challenging them 2 create safe conditions 4 Black ppl & marginalized groups on their platform? Fanning the flames of anti-Semitism resulting in a pipe bomb on George Soros' doorstep & campaigning against us using alt right media" (14 November 2018 at 8:36), online: *Twitter* <https://twitter.com/ColorOfChange/status/1062882090310664198> which included a link to Frenkel et al, *supra* note 200.

[208] See e.g., Sinduja Rangarajan, "Here's the clearest picture of Silicon Valley's diversity yet: It's bad. But some companies are doing less bad", *Reveal* (25 June 2018), online: <https://www.revealnews.org/article/heres-the-clearest-picture-of-silicon-valleys-diversity-yet/>; Sara Harrison, "Five Years of Tech Diversity Reports—and Little Progress", *Wired* (1 October 2019), online: <https://www.wired.com/story/five-years-tech-diversity-reports-little-progress/>.; Emily Reynolds, "Twitter's only black lead engineer quits over 'diversity issues'", *Wired* (4 November 2015), online: <https://www.wired.co.uk/article/black-twitter-engineer-quit-diversity>; and Sheelah Kolhatkar, "The Tech Industry's Gender-Discrimination Problem", *New Yorker* (13 November 2017), online: <https://www.newyorker.com/magazine/2017/11/20/the-tech-industrys-gender-discrimination-problem>.

[209] Laura Weiss, "Google workers, labor advocates confront parent Alphabet over practices", *Roll Call* (19 June 2019), online: <https://www.rollcall.com/2019/06/19/google-workers-labor-advocates-confront-parent-alphabet-over-practices/>. See also, e.g., "Despite Zuckerberg's reluctance to work with outsiders, experts probably could have warned him about human nature." Nitasha Tiku, "Facebook's Latest Fix for Fake News: Ask Users What They Trust", *Wired* (19 January 2018), online: <https://www.wired.com/story/facebooks-latest-fix-for-fake-news-ask-users-what-they-trust/> and @UntoNuggan, "WHO COULD HAVE PREDICTED that Twitter was a toxic cesspool of harassment? I'm sorry. Do you know zero trans women Like even online, do you not read their articles 🙃🙃🙃🙃🙃" (21 November 2017 at 13:50), online: *Twitter* <https://twitter.com/UntoNuggan/status/933044841579270151>.

subjects of focus.[210]

B. The Platform Foreseeability Gap

The discrepancy between what those who own and manage platform companies are ostensibly able to foresee, and what many platform employees, vulnerable and marginalized users, outside experts, academics, and human rights advocates have foreseen, gives rise to what one might call the *platform foreseeability gap.* The platform foreseeability gap consists of the range of platform affordances and reasonably foreseeable outcomes that are clear to a certain, usually marginalized, subset of users and experts, while being invisible to or dismissed as trivial, unlikely, and/or inconsequential by the platforms themselves. As a result, executives and management within platform companies—those with the power to implement solutions—fail to address serious issues through changes in policy, practice, business model, or technological design, leaving the way open for harmful consequences to later emerge. Adding injury to insult, such harm, by definition, is often more likely to befall the already marginalized groups who could have and in some cases in fact did warn against the precise kind of harm that occurred.

The rhetoric of "unintended consequences" in part sustains the platform foreseeability gap by emphasizing that negative impacts arising from online platforms are to be considered aberrations or "one-offs", when they may in fact be inherent features or the default slant of the platform (whether its own designers or owners realize it or not).[211] Those who design, manage, and own online platforms must more successfully parse between harmful consequences that are genuinely bugs and those that, more accurately, function as unacknowledged features. One way to do this is to distinguish between a particular platform's *intended* or *assumed* user base (e.g., well-meaning discussants engaging entirely in good faith while

[210] See e.g. "It's time for journalists to tell a new story that does not start the clock when privileged classes learn they are targets of surveillance. We need to understand that data has historically been overused to repress dissidence, monitor perceived criminality, and perpetually maintain an impoverished underclass. […] For targeted communities, there is little to no expectation of privacy from government or corporate surveillance. Instead, we are watched, either as criminals or as consumers. We do not expect policies to protect us. Instead, we've birthed a complex and coded culture—from jazz to spoken dialects—in order to navigate a world in which spying, from AT&T and Walmart to public benefits programs and beat cops on the block, is as much a part of our built environment as the streets covered in our blood." Malkia Amala Cyril, "Black America's State of Surveillance", *The Progressive* (30 March 3015), online: <https://progressive.org/magazine/black-america-s-state-surveillance-cyril/>; and "Viewing surveillance technologies through a framework of history and power helps us broaden mainstream concepts of what qualifies as surveillance. It *asks us to interrogate the capabilities of new devices and platforms—even those that may not have been originally designed with surveillance in mind*. For example, it could help us contextualize how smart home devices like Google's Alexa have provided a new avenue for spying on, and controlling, romantic partners—and the gendered nature of technology-enabled intimate partner violence. It can help us situate the fact that Grindr was sharing data on the HIV status and location of their users with other companies, and reflect on whether this kind of tracking could be linked to histories of queer surveillance. It brings to mind the growing use of 'e-carceration' practices; the GPS ankle monitoring of migrants and offenders as an 'alternative' to detention centres and prisons. It might help us speculate why Microsoft's commercial facial recognition technology had notoriously poor performance for people who weren't white, but conveniently improved its performance 'across all skin tones' shortly after its contract with ICE [Immigrations and Customs Enforcement, in the United States]—which involved developing technologies for the 'identification of immigrants'—was announced." Lorraine Chuen, "Watched and Not Seen: Tech, Power, and Dehumanization", *Guts* (3 December 2018), online: <gutsmagazine.ca/watched-and-not-seen/> (emphasis added).

[211] See e.g., "It has become clear that harassment is not an aberration but a condition of social media." Gillespie, *Custodians*, *supra* note 10 at 56.

marveling at how the 21st-century techno-utopia "brings the world closer"[212]), and the platform's *actual* user base (e.g., stalkers, doxxers, abusers, domestic terrorists), and assess whether a particular platform condition or affordance would, for that particular user base, be a significant drawback, or a useful feature.[213] Platform companies and their executives have long operated on a mental model that involves assuming the best of those who have the power to harm.[214] Meanwhile, their colleagues in adjacent fields (or adjacent offices within the same companies)—such as those developing facial recognition, algorithmic bail sentencing, "predictive policing" software, automated welfare decisions-making technologies, and artificial intelligence tools used in the immigration and refugee context—succeed by persuading decision-makers and the public to assume the worst of those with the least power in society and the most propensity to *be* harmed (including disproportionately criminalized Black and Indigenous communities, people of colour, LGBTQ+ communities, and those at socioeconomic disadvantage). [215]

Returning to the earlier example with Black feminist Twitter, "#EndFathersDay and the corrective rise of #YourSlipIsShowing is the story of a community that, mostly ignored by institutions, chose to fight back with the limited tools available."[216] Yet online platforms ignored both the abuse and calls for remedy,[217] and in return for their pains,

> Hudson, Crockett, and the other black feminists watched as the very same 4chan boards that had birthed #EndFathersDay spawned a new misogynistic harassment campaign mere months later: Gamergate. They watched as women like Zelda Williams and Zoë Quinn were aggressively bullied by accounts using many of the same tactics deployed during #EndFathersDay. And eventually, they watched as the 2016 election campaign unfolded and the very same forces that had been antagonizing them for years rebranded themselves as the alt-right.
>
> It wasn't just Twitter that seemed initially reluctant to take the problem seriously. In the years after #YourSlipIsShowing began, some media outlets diminished the danger of 'trolls' by characterizing

[212] Sarah Frier & Max Chafkin, "Zuckerberg's New Mission for Facebook: Bringing the World Closer", *Bloomberg* (22 June 2017), online: <https://www.bloomberg.com/news/articles/2017-06-22/zuckerberg-s-new-mission-for-facebook-bringing-the-world-closer>.

[213] Of course, this also involves recognizing if the user base consists of, more likely, all of these groups in addition to numerous others. The main point is that assuming the entire user base reflects only the first group has long been demonstrably erroneous, but many platform companies have been slow to change their policies, features, or practices accordingly.

[214] See e.g., Stephen Marche, "The Infuriating Innocence of Mark Zuckerberg", *New Yorker* (10 April 2018), online: <https://www.newyorker.com/culture/cultural-comment/the-infuriating-innocence-of-mark-zuckerberg>; and "When we built Superhuman, we focused only on the needs of our customers. We did not consider potential bad actors. I wholeheartedly apologize for not thinking through this more fully." Rahul Vohra, "Read Statuses" (3 July 2019), online: *Superhuman (on Medium)* <https://blog.superhuman.com/read-statuses-bdf0cc34b6a5>. (One missed consideration is that their customers would in all likelihood also include bad actors.)

[215] See generally, e.g., Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St Martin's Press, 2017); Andrew Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and The Future of Law Enforcement* (New York: New York University, 2017); Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Cambridge: Polity Press, 2019); Erin Corbett, "Tech Companies Are Profiting Off ICE Deportations, Report Shows", *Fortune* (23 October 2018), online: <https://fortune.com/2018/10/23/tech-companies-surveillance-ice-immigrants/>; and Empower LLC, *Who's Behind ICE? The Tech and Data Companies Fueling Deportations* (2018), online: *Mijente* <https://mijente.net/wp-content/uploads/2018/10/WHO%E2%80%99S-BEHIND-ICE_-The-Tech-and-Data-Companies-Fueling-Deportations_v3-.pdf>.

[216] Rachelle Hampton, "The Black Feminists Who Saw the Alt-Right Threat Coming", *Slate* (23 April 2019), online: <https://slate.com/technology/2019/04/black-feminists-alt-right-twitter-gamergate.html>.

[217] "But despite the evidence that harassment campaigns fueled by a noxious mixture of misogyny and racism spelled out a threat to users from vulnerable groups, [activists] Hudson and Crockett felt that Twitter basically did nothing." *Ibid.*

> their flirtation with white nationalism as tongue-in-cheek—until those trolls took their rhetoric offline and onto the streets of Charlottesville, Virginia. […] [M]ultiple journalists described feeling like they'd made light of the rhetoric they'd seen online, operating under an assumption that real, violent hatred and ironic provocation were separate worlds. […]
>
> But to Hudson and Crockett and the other black feminists behind #YourSlipIsShowing, the idea that violent speech on the internet could easily translate to an actual physical threat was never a question. "If your body … has never been subject to this kind of racist violence, if you have never been under threat because of how you look or who you are in the world, then you're going to be less inclined to take attacks against other people's bodies very seriously," said Phillips. "Because for you your body has only been a vessel of safety … those women [who created #YourSlipIsShowing] understood in an embodied, visceral way that this is not playful trolling … the majority of people didn't pay attention." [218]

That aggressive, sustained, coordinated, and targeted racist and misogynistic abuse for some is "ironic provocation" at most, while for others its hairline separation from physical violence "was never a question", represents a platform foreseeability gap within which those who belong to marginalized communities live and die.

C. Unstable Nature of Emergence in "Emergent Systemic Harm"

In light of the demonstrable distinctions between "unintended", "unforeseen", and "unforeseeable", the use of "emergent" as applied to platform-facilitated systemic harms becomes more fragile than it was in the first half of this paper. There is also acknowledged tension inherent in the idea that a systemic harm can be both truly emergent and yet sufficiently foreseeable so as to constitute grounds for liability in negligence (discussed in Part IV below). This tension is due to the platform foreseeability gap. The notion of emergent systemic harm arose from the seemingly oft taken-for-granted idea that systemic harms arising from online platforms were "unintended consequences", with the implicit notion that such consequences were thus equally unforeseen. Upon closer examination, however, several of these systemic harms were only "unforeseen" by and thus "emergent" to the platform companies themselves; to outsider observers such as vulnerable users who experience systemic discrimination, marginalized employees, human rights advocates, and academics, as well as malicious actors, many of these same systemic harms or the risk of them were absolutely foreseen, demonstrating they were not strictly emergent after all. One example that illuminates this difference is platform developers' general inability to foresee the existence of bad faith or abusive users exploiting their platform features to cause harm. Without that node existing in one's mental model of the platform system, resulting harms to marginalized and vulnerable users may appear "unexpected" and "emergent", whereas with that node included in one's mental model, such consequences are reasonably foreseeable, if not obvious, from the beginning.

At the same time, some situations remain where a platform-facilitated systemic harm can indeed be both emergent, in the strict sense of the definition provided,[219] *and* subject to the platform foreseeability gap (i.e., foreseen by some but not by the platform). An example would be the negative impacts of Uber, Lyft, and Airbnb in municipalities where their forms of ridesharing services and short-term rental accommodation services, respectively, were not illegal and did not violate existing bylaws. The lack of legal prohibition meant that none of the individual components of each platform ecosystem (the apps, services, drivers, passengers, guests, and hosts) were necessarily operating illegally, unethically, or unreasonably. However, Uber and Lyft still gave rise to the emergent systemic harm of impairing public

[218] *Ibid*.

[219] Emergent systemic harm refers to a situation where there is an integrated system in which none of the components or actors' conduct were individually illegal, ill-intentioned, unreasonable, or unethical, but which results in all the components and actors working together in a way that gives rise to systemic harm.

transit ridership and increasing traffic congestion, while Airbnb devastated local rental markets and affordable housing crises. These consequences were emergent according to the definition in this paper, but it is possible or even likely that transit and road policy experts, housing economics experts, tenants' rights advocates, and those dependent on public transit could have foreseen the respective risks, without the platform companies themselves foreseeing anything.

The remainder of this paper will continue predominantly to use the term "emergent systemic harm" for consistency and ease of reference, but with occasional uses of "quasi-emergent" or "pseudo-emergent" to reflect the unstable nature of the characteristic. However, going forward, emergent systemic harm should be understood to mean systemic harm that appears emergent, or that a given party considers emergent—and indeed may turn out to have been truly emergent—but that upon further investigation or disclosure of new information, may turn out at a later point to not have been emergent after all.

In light of the unstable nature of the "emergence" in emergent systemic harm, a key purpose of the analysis of reasonable foreseeability in Part IV below, as applied to online platform-facilitated systemic harms, is to develop a more comprehensive interpretation of reasonable foreseeability in the legal test for negligence. This approach represents an attempt to use law to close the platform foreseeability gap, by encouraging platform companies to develop greater foresight, at the level of those who are vulnerable to particular risks of systemic harm as a result of the companies' platform systems.

Part IV. Platform-Facilitated Emergent Systemic Harm and Reasonable Foreseeability

The implied logical conclusion from discourse that centres on the "unintended consequences" of platform-facilitated emergent systemic harms is that because they did not mean to do it, the platform companies or their owners and executives cannot or ought not to be held legally liable for the harms they contributed to (and indirectly profited from). However, Canadian law does provide for holding parties accountable for harm even in the absence of intent, including in the areas of human rights law, environmental law, and certain categories of tort law. Part IV will specifically examine the foreseeability component of negligence claims in tort law, as applied to platform algorithm-facilitated emergent systemic harm to marginalized groups.

This paper focuses solely on foreseeability in the hopes of contributing to the duty of care, standard of care, and causation analyses in the future, to the extent foreseeability is in issue under any of these elements in a given case, and to prevent the rhetoric and associated implications of "unintended consequences" from resulting in legal claims around platform-facilitated emergent systemic harms being rejected out of hand.[220] For reasons of time and length, the paper does not purport to analyze duty of care, standard of care, or remoteness in their own rights, with respect to platform-facilitated emergent systemic harm. However, such an exercise would be valuable to undertake in future scholarship, in addition to analysis evaluating how other elements of the test for negligence would apply to online platforms for emergent systemic harm, including those elements that do not directly include foreseeability, such as proximity (at the duty of care stage), causation, and possible defenses other than remoteness.

[220] This is not to say that such claims may not be rejected on the basis of other elements of the negligence analysis, or for other reasons outside of tort law—section 230 of the *Communications Decency Act* in the United States comes to mind, which does not yet have an implemented equivalent in Canada, though a provision in USMCA waits in the wings—but again, unfortunately, analysis of such contentions must be left for another day and another paper.

A. Tort Law and Applying Negligence to Platform-Facilitated Emergent Systemic Harm

i. Tort Liability as an Avenue of Recourse

Tort law is proposed as only one additional avenue of recourse for marginalized platform users (and potentially non-users) who have suffered damage as a result of platform-facilitated emergent systemic harm, providing a private right of action where regulation has failed to prevent harm. There is no intention to replace or argue against pre-existing legal paths such as bringing a complaint under human rights legislation. Rather, tort law offers certain advantages that may not be available under statutory human rights regimes, and could supplement their deficiencies while providing more options to those who are in need of remedy and redress. For instance, tort law would allow for collective action through class proceedings, where damages are too small or are non-monetary and not sufficiently conducive to quantifying to justify an individual bringing claim, but where a systemic issue has adverse consequences for a particular marginalized group across the board. At the same time, where monetary damages can make a difference, higher amounts are available through tort actions compared to human rights legislation[221]—meaning they may both better serve those harmed by platform-facilitated emergent systemic harms while also deterring negligence more effectively rather than being absorbed as a small cost of doing business. Tort law also provides an important normative force as "an ethical system promoting the personal responsibility of wrongdoers for the harm that they have caused to others"[222] and a mechanism through which the vulnerable and the marginalized can hold the powerful to account and seek redress for harms from power imbalances:

> Drawing from empowerment theories, Shuman remarks that "tort litigants may enlist the coercive power of the judicial system to reshape the power imbalance in their relationships" with defendants. This potential benefit of the tort system is particularly important in the context of racial discrimination [and other forms of systemic discrimination against marginalized groups], where issues of power and dominance are central to the relationship between plaintiff and defendant.[223]

Finally, relying on a robust interpretation of proximity (under the duty of care analysis) in today's connected world, tort law may allow impacted marginalized groups, or injured individuals belonging to such groups, to bring action against an online platform for emergent systemic harm even if they were not users of the platform. No underlying contractual relationship would be required to bring a claim, thus also preventing potential litigants from being bound by mandatory arbitration clauses or other provisions that

[221] See e.g. "It is noteworthy that the total award for pain and suffering, $10,000, all flowing from the exploitation of the plaintiff's disability, was twice as much as the cap permitted under the *Canadian Human Rights Act* at the time. Nor would the additional $10,000 for deterrence, indicating society's censure against such behaviour, have been allowable under the *Act*. This is a case where the public system's unique remedial power to reinstate would have been useless to Ms. Boothman, given her continued inability to work, as a result of the trauma…" Tamar Witelson, "Retort: Revisiting *Bhadauria* and the Supreme Court's Rejection of a Tort of Discrimination" (1999) 10:2 NJCL 149 at 179; and Philip H Osborne, *The Law of Torts*, 5th ed (Toronto, ON: Irwin Law Inc, 2015) at 289.

[222] Osborne, *supra* note 220 at 12; see also: "[T]he critical role tort law plays in regulating social conduct … tort law makes definitive determinations of what is permissible and impermissible within a community. Such determinations depend on theories about morality — what sort of conduct is acceptable or wrongful — and equally about theories of social relationship — how the words or acts of one person affect the well-being of another." Rakhi Ruparelia, "'I Didn't Mean It That Way!': Racial Discrimination as Negligence" (2009) 44 SCLR (2d) 81 at 85.

[223] Ruparelia, *supra* note 221 at 86-87 (footnotes omitted); See also: "The role of tort law as ombudsperson has been identified and promoted by Mr. Justice Linden.43 He has pointed out that tort law is well placed to challenge the wrongful and harmful behaviour of the most powerful persons and institutions in Canada. From time to time, concern is expressed about the accountability of the rich and powerful in society and the effectiveness of public law controls on their activities. In these situations, the accountability and sanctions of tort law may be advantageous because tort litigation can be initiated by private individuals and it is adjudicated by judges who are independent of political control." Osborne, *supra* note 220 at 17

deter access to justice, within a platform's Terms of Use or End-User License Agreement. The availability of legal recourse to non-users who are nonetheless impacted by a platform's activities (due to belonging to a marginalized group who is disproportionately harmed) is particularly important in light of digital platforms' now well established roles as monopolistic arbiters of public discourse, public life, and public policy, to the extent the latter is shaped by what is made possible or not possible for users to do on a given platform. Gillespie's following description of how on-platform activities can equally impact non-users and the wider public helps to illuminate why it is important that members of marginalized communities impacted by platform-facilitated harms, who themselves are not users of that platform, also have standing to bring legal action:

> Even if I never saw, clicked on, or liked a fraudulent news post, I still worry others may have done so. I am troubled by the very fact of it and concerned for the sanctity of the political process as a result. Protecting users is no longer enough. The offense and harm in question is not just to individuals but also to the public itself and to the institutions on which it depends. This, according to John Dewey, is the very nature of a public: "The public consists of all those who are affected by the indirect consequences of transactions to such an extent that it is deemed necessary to have those consequences systematically cared for." What makes something a concern to the public is the potential need for its inhibition. [224]

That non-users of a given online platform are able to seek remedy against that platform is particularly important where they have no control over any of the components in a platform ecosystem that gave rise to emergent systemic harm, including other platform users engaging in activities that immediately concern the non-user, but without their consent. Examples of this situation would include non-consensual distribution of intimate images, someone getting doxed on a platform they do not use, someone being impersonated on a platform they do not use, or the disclosure of their personal information through a platform they do not use, to third parties or the platform itself.[225]

ii. The Tort of Negligent Discrimination

The focus of this paper is emergent systemic harm to marginalized groups in particular, which raises the spectre of the tort of discrimination, the existence of which was rejected and foreclosed, at least under some circumstances, by the Supreme Court of Canada in *Seneca College v. Bhadauria*.[226] However, some scholars and members of the judiciary in lower courts have provided critiques, rebuttals, or workarounds in certain decisions distinguishing *Bhadauria*,[227] suggesting that the decision may eventually be

[224] Gillespie, "Platforms", *supra* note 5 (footnotes omitted). See also *Facebook agrees to stop using non-users' personal information in users' address books* (23 May 2018), PIPEDA Report of Findings #2018-003, online: *Office of the Privacy Commissioner of Canada* <priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2018/pipeda-2018-003/> [OPC, *Non-Users' Personal Information*].

[225] See e.g., OPC, *Non-Users' Personal Information*, *supra* note 223; and *In response to a case of a teen who was a victim of online impersonation, Facebook agrees to help non-users, on a case-by-case basis, reinstate their on-line reputation*, PIPEDA Report of Findings #2013-010 (11 July 2013), online: *Office of the Privacy Commissioner of Canada* <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2013/pipeda-2013-010/>.

[226] *Seneca College v Bhadauria*, [1981] 2 SCR 181 [*Bhadauria*].

[227] See e.g. "The recent obiter dictum of LeBel J. in *Keays* hinted at a readiness to reconsider *Bhadauria*. He stated that it was not necessary for the Supreme Court in *Bhadauria* 'to preclude all common law actions based on all forms of discriminatory conduct'. Furthermore, he observed the change in the legal landscape and suggested that tort law should not remain static. We should seize this opportunity to forge a critical role for tort law in remedying discrimination." Ruparelia, *supra* note 221 at 105 (footnotes omitted); and "[T]he [Manitoba] Court of Queen's Bench faced the issue squarely and ruled that a new kind of tort based on discrimination does exist: 'In my opinion in today's world and under the circumstances of this case sexual harassment should constitute a wrong or tort. The remedy then should be damages.' Joanne Lajoie, a seventeen-year-old single mother, worked for less than a month as a part-time waitress for Michael Kelly. During that time he made sexual comments and advances to her, until she

overturned or at least narrowed in influence and relegated to a small corner of the law[228]—more so if societal and legal norms with respect to systemic discrimination and equality rights advance further. Additionally, some contend that even assuming *Bhadauria* is here to stay, human rights legislative regimes leave room for actions in tort where they are inadequate to address a particular issue or ground, including if the plaintiff belongs to a marginalized community defined by a characteristic not specifically enumerated in the governing statute.[229]

More pertinently, *Bhadauria* may not pose a barrier to the analysis proposed in this paper because the decision concerns an intentional tort of discrimination, whereas by definition, platform-facilitated emergent systemic harm to a marginalized group would have to be an *unintentional* tort.[230] At this point, the analysis dovetails with what Rakhi Ruparelia has proposed as the tort of negligent discrimination. While Ruparelia developed this tort in the context of racial discrimination specifically, her framework and reasoning can apply to discrimination on the basis of other protected or marginalized characteristics, particularly where they also involve a history of systemic oppression that evolves and continues through present day.[231] According to Ruparelia, compared to intentional torts, "Negligence is better situated to

quit. The Court found that it would have been most appropriate for Ms. Lajoie 'to lay a complaint ... under the [Manitoba Human Rights] *Code* ... However ... the *Code* is not given exclusive jurisdiction.' In support of this position, the Court cited Prosser's view that the law of torts is not static and the limits of its development are never set." Witelson, *supra* note 220 at 181

[228] See e.g., "When the Supreme Court concluded that Pushpa *Bhadauria* could not pursue her claim of discrimination in the civil courts, that general bar to an action outside the human rights system could have been read in at least three ways: ... 2) a "middle road" approach is that only a general tort of discrimination is barred, but not discrimination which can be framed as an already recognized cause of action; and 3) the most creative reading is that although a general tort of discrimination is barred, that does not preclude traditional private actions to incorporate a discrimination issue, nor does it prevent the development of new, specific types of actions aimed at discriminatory conduct. As the caselaw has evolved in the wake of Bhadauria, it appears that the middle road approach has won favour, and the most creative reading is also starting to emerge. Taken together, these modern applications of *Bhadauria* have the potential to eclipse the case's significant barrier to private claims for redress from the harm of discriminatory acts." Witelson, *supra* note 220 at 175.

[229] See e.g. "In short, it is argued that where a human rights code is silent with respect to a ground of alleged discrimination, and when a specific relationship which could be identified as proximate is at issue, a consistent reading of *Bhadauria* would lead to the conclusion that neither access to the common law nor the civil law is a priori foreclosed. In such circumstances, courts of general jurisdiction are authorized to determine whether an exercise of one's legal powers in such a discriminatory manner is consistent with legal doctrines and principles governing the use of such legal power." Amnon Reichman, "Professional Status and the Freedom to Contract: Toward a Common Law Duty of Non-Discrimination" (2001) 14:1 Can JL Juris 79 at 88-89 (footnotes omitted); see also "[B]ased on the Court's view, a close reading of the *Act*, and an analysis of the disposition of complaints received by the Commission, there is evidence that complainants may not receive a full and fair opportunity to seek redress when they feel they have suffered discrimination prohibited under the federal human rights system. None of the officials at the administrative stage is required to have formal legal training. There is no appeal of decisions by intake officers or investigators, and limited judicial review of Commission decisions to refuse sending a complaint to tribunal. Time and budgetary constraints may influence all of these procedures. Finally, some or all of these weaknesses in the system may account for the very small proportion of complaints that make it to the adjudicative stage. All of these elements of the federal human rights system support a legitimate concern that the operation of the system may prevent a complainant from ever receiving his or her 'day in court'." Witelson, *supra* note 220 at 167.

[230] The cause of action does not necessarily have to be a tort of discrimination. For example, the claim could be framed as applying equality and non-discrimination principles to the analysis within a pre-existing tort, as Witelson, *supra* note 220 at describes parties and courts have done in the wake of *Bhadauria*.

[231] "My paper begins with a brief discussion on why tort law is an important venue in which to deal with racial discrimination. Most of my arguments will be applicable to other forms of discrimination as well. However, this paper is concerned primarily with racial discrimination because of the distinctive socio-legal history of racism and the unique and fluid ways in which it has been, and continues to be, manifested. 12 In my view, these

respond to the reality that racism is not a series of discrete actions but rather 'an integrated system that elevates one group at the expense of another'."[232]

Negligence not only aligns with the focus on impact regardless of intent, as required in human rights law, but Ruparelia's description of negligence also fits situations that give rise to platform-facilitated emergent systemic harms to marginalized groups: such harms reflect the reality of adverse effects discrimination across online platforms, arising not from overt actions, but from a system of integrated components giving rise to conditions that disproportionately harm a given marginalized group in particular, if not exclusively. For example, members of certain racialized communities, such as Black and Indigenous individuals, and racialized religious communities, such as Muslim individuals, are particularly put at physical and psychological risk as a result of YouTube's ad-driven and "engagement-optimizing" recommendation engine facilitating the sustained elevation and circulation of extremist white supremacist videos. Similarly, Safiya Noble has demonstrated how online searchers' biases combined with Google's "neutral" search algorithms have combined to the specific detriment of Black women and girls, through top-ranked search results promoting sexual objectification, with the implicit notion that pornography is the most salient information possible, out of the entire Internet, relevant to Black women and girls.[233]

Further in favour of applying negligence, the Supreme Court of Canada has recognized the doctrine of *systemic negligence*, albeit thus far predominantly in the context of class action proceedings against "fiduciary-like institutional settings (primarily residential schools)".[234] Systemic negligence is "negligence not specific to any one victim but rather to the class of victims as a group",[235] and is found based on the defendant's "failure to have in place management and operations procedures that would reasonably have prevented the abuse".[236] One could theoretically apply this standard to evaluate a platform company's policies, processes, and practices with respect to designing and managing their platform and how its algorithms and affordances interact with known or likely user behaviour. Moreover, a recent case in Ontario expanded the notion of systemic negligence to explicitly include both omissions and actions:

> In *Rumley v. British Columbia*, the Supreme Court had a negative definition of systemic negligence, which is to say a definition based on acts of omission. The systemic negligence of the province was its failure to have in place procedures that would have prevented the tortious conduct. But a positive definition of a systemic negligence, which is to say a definition based on actions of commission, would have been *carelessly having in place harmful procedures and practices* contrary to the intent of the enabling legislation.[237]

With respect to "carelessly having in place harmful procedures and practices", one cannot help but think of above-mentioned companies like Snapchat and Superhuman boasting of granular real-time location-tracking features in their products, without having ostensibly benefited from first consulting any women

considerations, combined with the particularly biting stigma attached to an accusation of racism, warrant a separate discussion on how tort law should redress racial discrimination." Ruparelia, *supra* note 221 at 84

[232] Ruparelia, *supra* note 221 at 95.

[233] Noble, *supra* note 7.

[234] Margaret Isabel Hall & Aliya Chouinard, "Systemic Wrongdoing, Public Authority Liability, and the Explanatory Function of Tort Doctrine: Two Case Studies" (2018) 84 SCLR (2d) 71 at 82.

[235] *Rumley v British Columbia*, 2001 SCC 69, at para 34 [*Rumley*]. See also *Tippett v Canada*, 2019 FC 869, at paras 64-65.

[236] *Rumley*, *supra* note 234, at para 30.

[237] *Reddock v Canada (Attorney General)*, 2019 ONSC 5053, at para 426 (emphasis added). The judge continues, "While not a systemic negligence case, *Little Sisters Book and Art Emporium v. Canada (Minister of Justice)*, mentioned several times above, demonstrates how a systemic breach of a citizen's *Charter* rights is not limited to acts of omission; i.e., failing to introduce safeguards, but includes mismanaging a system, acts of commission." *Ibid*, at para 429.

or members of other marginalized groups historically subjected to invasive tracking and surveillance-enabled abuse, whether by intimate partners, other platform users, private companies, or law enforcement.

The next section of this paper will examine how reasonable foreseeability is traditionally determined in the negligence analysis for unintentional torts, then provide a critical intersectional recalibration of how foreseeability might be interpreted to update the concept for contemporary understanding of systemic discrimination and inequality, in the context of online platforms and emergent systemic harms to marginalized groups. The analysis will not turn on negligent discrimination or any one tort specifically, but will focus on how the foreseeability analysis ought to be applied in principle, regardless of the specific tort at hand or the element of the tort under discussion. It should not thus matter, for example, whether the context is assessing whether there is reasonable foreseeability for the sake of determining if a duty of care exists under a novel tort, or evaluating whether a harmful outcome was not reasonably foreseeable so as to establish remoteness in the context of a statutory tort. The focus is on reasonable foreseeability itself, and the analysis is meant to apply in all contexts where foreseeability is in contention, with respect to tort claims involving platform-facilitated emergent systemic harm.

B. Applying Reasonable Foreseeability to Platform-Facilitated Emergent Systemic Harms

Reasonable foreseeability is a core concept that warrants particular examination due to the central role it plays in three core elements of determining tort liability based on negligence: duty of care, standard of care, and remoteness. First, to establish whether a duty of care exists for the defendant to have breached, "the plaintiff must provide a sufficient factual basis to establish *that the harm was a reasonably foreseeable consequence of the defendant's conduct* in the context of a proximate relationship."[238] Second, reasonably foreseeable risk is the "central element in applying the standard of reasonable care",[239] in determining whether or not a defendant acted with sufficient care to ward off liability for negligence, or acted with sufficient carelessness so as to attract liability. As Osborne states, "Conduct is negligent only if it carries a risk of damage that a reasonable person would contemplate and guard against."[240] Third, determining reasonable foreseeability is necessary to analyzing the element of remoteness, which is the principle that even if the plaintiff's conduct technically did cause the harmful consequence, the harm was "too remote to the wrongful conduct to hold the Defendant fairly liable."[241]

Assessing whether a particular harm, or risk of harm, was reasonably foreseeable involves a range of considerations, seven of which will be discussed in the following subsections. This is not an exhaustive list but a selection of key factors that appeared throughout leading jurisprudence on reasonable foreseeability in negligence, and which are particularly salient to determining reasonable foreseeability in the context of platform-facilitated emergent systemic harms.

[238] *Rankin (Rankin's Garage & Sales) v JJ*, 2018 SCC 19, at para 19 (emphasis added) [*Rankin's Garage*]. Duty of care is a longstanding foundational cornerstone of Canadian tort law: "If it is necessary to determine whether a novel duty exists, the first stage of the *Anns/Cooper* test asks whether there is a relationship of proximity in which the failure to take reasonable care might foreseeably cause loss or harm to the plaintiff… Once foreseeability and proximity are made out, a prima facie duty of care is established." *Ibid* at para 18 (inline citations omitted).

[239] Osborne, *supra* note 220 at 30.

[240] Osborne, *supra* note 220 at 30.

[241] *Becker v City of Toronto*, 2019 ONSC 3912, at para 209 [*Becker*]. In other words, remoteness is what characterizes "situations where the loss is so different from what one might have expected, so disproportionate to the magnitude of the fault, or so fluky or bizarre that it is unfair to hold the defendant legally responsible for it." Osborne, *supra* note 220 at 98.

i. Type of Damage and Class of Plaintiff

First, reasonable foreseeability requires demonstrating that "the risk of the type of damage that occurred was reasonably foreseeable to the class of plaintiff that was damaged."[242] In this case, the "class of plaintiff" would be a particular marginalized group, defined by a characteristic protected under section 15 of the *Charter*, listed under human rights statutes, or on the basis of social science evidence accepted by the court. The nature of the group may influence how reasonably foreseeable the type of damage that occurred to them was. For example, it would likely be found reasonably foreseeable that female journalists or female politicians on a social media platform would sustain damage falling into the general category of online abuse, in the form of death and rape threats, hate speech, and ongoing harassment. A class of victims/survivors of intimate partner abuse and stalking could establish that it is reasonably foreseeable that they would be subjected to further abuse and stalking as a result of a platform's real-time location-tracking function. A group of tenants who were evicted and had evidence their former units were turned into Airbnb rentals could potentially bring action against the platform based on the foreseeable harm of lost housing from no-fault eviction, to the detriment of low-income tenants in particular.

ii. "Precise Concatenation of Events"

Critically, the test for reasonable foreseeability does not require that the specific form and extent of damage that in fact occurred was foreseeable, only the "type" of damage. This overlaps with the second consideration: reasonable foreseeability does not require the ability to foresee "the manner in which the accident occurred, the mechanics of the accident",[243] or "the precise concatenation of events that led to the harm".[244] Moreover, the "extent of the damage and its manner of incidence need not be foreseeable if physical damage of the kind which in fact ensues is foreseeable."[245] This principle, provided it applies equally to non-physical damage, is particularly useful with respect to platform-facilitated emergent systemic harms in light of some of the more innovative strategies that ill-intentioned actors have used in exploiting platform affordances to harm others.[246] The high-level nature of what must be reasonably foreseeable may also assist negligent discrimination claims in the sense that the specific route by which a historically marginalized group experienced adverse impact discrimination across a platform ecosystem may not have necessarily been foreseeable, but the discrimination itself may have been. This principle also particularly applies to determining platform liability for emergent systemic harm, because it requires being able to foresee the general category of harm that occurred, or the risk of it, but does not require being able to foresee exactly how different components of the platform's system interacted with each other to produce the specific type of harm that occurred.

iii. Likelihood of Harm

The third consideration in determining reasonable foreseeability is assessing the likelihood of the damage that occurred: how probable was the harm? The Supreme Court of Canada described the necessary threshold as follows:

[242] *Rankin's Garage*, *supra* note 237 at para 24.
[243] Osborne, *supra* note 220 at 100
[244] *R v Côté et al*, [1976] 1 SCR 595, 51 DLR (3d) 244, 1974 CanLII 31 (SCC).
[245] *School Division of Assiniboine South, No 3 and Hoffer et al v Greater Winnipeg Gas Company Limited*, [1971] 4 W.W.R. 746 (Man. C.A.), 21 DLR (3d) 608, 1971 CanLII 959 (MB CA) (leave to appeal dismissed, 1 March 2001) [*Assiniboine*].
[246] See e.g., the coordinated flagging campaigns described in Crawford & Gillespie, *supra* note 84.

> The degree of probability that would satisfy the reasonable foreseeability requirement was described in *The Wagon Mound (No. 2)* as a "real risk", i.e. "one which would occur to the mind of a reasonable man in the position of the defendan[t] . . . and *which he would not brush aside as far-fetched*".[247]

Immediately the problem and dangers of the platform foreseeability gap become clear as a roadblock to equity, were it to be enshrined in law through standards of what is considered reasonably foreseeable, and what is not. Consider, for example, the culminating violence and threats to physical safety that Twitter and non-marginalized journalists "brushed aside as far-fetched", rather than took seriously as a reasonably foreseeable consequence of the online abuse hurled by the self-styled "alt right" at Black, Indigenous, and racialized women and men, in addition to members of other communities facing systemic discrimination. While "the reasonable man" has since been updated to "the reasonable person" in legal writing, feminist legal critique asserts that this change in terminology does not necessarily reflect underlying change in values or perspectives, such that "to continue to insist upon the reasonable person now as being gender neutral will result in the perpetuation of the male bias",[248] but a bias now concealed and made harder to acknowledge and account for. The concept of the reasonable person will be revisited as part of the more comprehensive reinterpretation of foreseeability advanced below.

iv. Chain of Foreseeability

The fourth consideration in reasonable foreseeability is the ability to meet the test through constructing a "foreseeability chain" linking the defendant's conduct to the resulting harm. This involves "divid[ing] the causal sequence into a number of discrete steps, each of which is a readily foreseeable consequence of the preceding step",[249] with *Assiniboine South School Division No. 3 v. Greater Winnipeg Gas Co.* demonstrating how far this exercise can go.[250] A similar chain of foreseeable consequences might give rise to establishing reasonable foreseeability in the case of YouTube's recommendation algorithm "building a dystopia", the components of which are all individually foreseeable. That is to say, it was foreseeable that YouTube would want to maximize viewer engagement; that an algorithm optimizing for engagement would use proxy metrics such as view time; that engagement metrics are particularly high for sensational, outrageous, or shock-value content; that users would want to produce videos that maximize engagement (particular in view of YouTube actively providing them with incentives); that the recommendation engine would likely promote videos it deems likely to be the most engaging; that all else being equal, people on YouTube would be more likely to watch videos that were recommended to them,

[247] *Mustapha v Culligan of Canada Ltd*, 2008 SCC 27, at para 13 (in-line citations omitted, emphasis added) [*Mustapha*]; *Becker*, *supra* note 240 at para 210.

[248] Sherilyn J Pickering, "Feminism and Tort Law: Scholarship and Practice" (2010) 29 Windsor Rev Legal Soc Issues 227 at 239. Pickering also states, "[M]any feminist scholars refuse to identify the reasonable person as such, but rather identify the so-called abstract, universal person as the reasonable man. In changing the phrase from the reasonable man to the reasonable person, the court did not actually change the content or the character of the reasonability standard but rather continues to demand that the person be reasonable by the old standards of reasonability and to invite judicial decisions based upon a hidden set of gendered values." *Ibid* at 239 (footnotes omitted).

[249] Osborne, *supra* note 220 at 102.

[250] "The [foreseeability linkage / chain] technique was used in *Assiniboine South School Division No. 3 v. Greater Winnipeg Gas Co*. In that case, the defendants' failure to start a snowmobile with reasonable care resulted in fire damage to the plaintiff's school. The risk inherent in the starting procedure was that the snowmobile might take off without its rider to the peril of persons and property in the vicinity. The chain of events that led to the fire was broken down into a series of foreseeable occurrences. They included the foreseeability of impact with a building, foreseeability of gas-riser pipes on buildings in that area of Winnipeg, foreseeability of impact with a gas-riser pipe, foreseeability of the escape of gas from the impact with a pipe, and foreseeability that gas might find its way into the school where it might be ignited by a foreseeable pilot flame in the boiler room. Foreseeability was thereby established and liability was imposed. This technique of building foreseeability on foreseeability is not uncommon in remoteness cases." Osborne, *supra* note 220 at 102-03.

than videos that were not recommended; that likeminded content creators may unite and collaborate; and that online polemic can lead to real-world consequences embodying such polemic. Each of these links on its own is a reasonably foreseeable consequence of a prior decision, such that an *Assiniboine* approach could result in building atop them all to find negligence on YouTube's part in designing and upholding a system that was on course to produce Tufekci's dystopia, including its impacts on marginalized users.[251]

v. Inaction Despite Actual Knowledge

The fifth consideration is that courts may impose liability for negligence where there is evidence that the defendant had knowledge they could have acted on to prevent the harm that occurred, or mitigate the risk that it would occur, but did not in fact act on that knowledge. In *Q et al. v. Minto Management Ltd. et al* ("*Minto Management*"), for example, a tenant was sexually assaulted and raped by her landlord's employee, who had likely entered her apartment using a master key.[252] The court found the landlord negligent for having "failed to take any reasonable measures for the protection of the tenant from the risk of a repetition of the May incident [where another woman in the same building had been raped three months prior, by the same employee] or an incident of a similar kind."[253] The landlord did not, for instance, upgrade tenants' apartment locks or do anything about the fact that "[m]aster keys were available to many employees",[254] despite the fact that the landlord "knew about the earlier rape and knew it likely had been committed by someone with a master key",[255] having been advised by the police that they thought the May perpetrator was an "insider".[256] Further, the court held that the landlord did not have to foresee which specific employee broke in and attacked the tenant, but should have reasonably foreseen unauthorized entry and related criminal activity enabled by it.[257] It is not difficult to see how this principle could apply to platform-facilitated emergent systemic harm in cases where, for example, platform companies were warned of potential or even certain harms down the line by their own employees, yet ignored or were dismissive of them (both the warnings and the employees).[258]

vi. History and Precedents as Evidence

Sixth, evidence that contributing factors of the harm were reasonably foreseeable may be required to find that the harm itself was reasonably foreseeable. For example, in *Rankin (Rankin's Garage & Sales) v. J.J.*, 2018 SCC 19 ("*Rankin's Garage*"), the Supreme Court of Canada held, "Some evidentiary basis is required before a court can conclude that the risk of [car] theft [from the defendant's premises] includes the risk of theft by *minors*."[259] The foreseeability of theft by minors was central to the analysis because the court had to determine if the garage owners were liable for the injuries of the minors who stole an unlocked vehicle, with keys left inside, from the premises. In finding no reasonable foreseeability, the majority stated, "Rankin's Garage had been in operation for many years and no evidence was presented to suggest that there was ever a risk of theft by minors at any point in its history."[260] The majority contrasted the facts before them to an earlier case, *Holian v. United Grain Growers Ltd.*, where "the defendant's

[251] See e.g., Danny Nett, "Is YouTube Doing Enough To Stop Harassment Of LGBTQ Content Creators?" *NPR* (8 June 2019), online: <https://www.npr.org/2019/06/08/730608664/is-youtube-doing-enough-to-stop-harassment-of-lgbtq-content-creators>.
[252] *Q et al v Minto Management Ltd et al*, 49 OR (2d) 531 (ONSC), 15 DLR (4th) 581, 1985 CanLII 2103 [*Minto Management*] (aff'd by ONCA in *Q et al v Minto Management Ltd et al*, 57 OR (2d) 781, 34 DLR (4th) 767).
[253] *Ibid*.
[254] *Ibid*.
[255] *Ibid*.
[256] *Ibid*.
[257] *Ibid*.
[258] See e.g., Bergen, *supra* note 121.
[259] *Rankin's Garage*, *supra* note 237 at para 46 (emphasis in original).
[260] *Ibid* at para 66.

employees knew that children used the area near the storage shed as a shortcut. This made it reasonably foreseeable that minors may have stolen from the storage shed."[261] In the context of online platforms, the existence of "unintended consequences", or emergent systemic harms, having occurred in the past should serve as an evidentiary basis that future similar harms from the same platform are reasonably foreseeable. For instance, within Facebook's fourteen years[262] and counting of user outcries, public and political censure, and regulatory threats and admonishment over non-consensual data collection and sharing, third-party data breaches and technically authorized data scraping, and thoughtless failures around privacy control, such as making certain kinds of content public by default without user consent or notice—it seems fair to hazard that after the first 1-2 years, let alone subsequent 5, 7, or 10 years, similar privacy harms from Facebook's platform ecosystem should be considered reasonably foreseeable. On this basis, if the other elements of the tort were met, Facebook might be found liable for systemic negligence giving rise to privacy-related harms to users across the board, with disproportionate impact on marginalized users who more greatly rely on privacy measures for physical and psychological safety.

vii. New Intervening Act

The seventh consideration is the possible defence of *novus actus interveniens*, or a "new intervening act" by a third party which breaks the chain of causation between the defendant's conduct and the resulting harm.[263] This consideration is particularly significant in the context of online platform-facilitated systemic harms that are not necessarily emergent, where the damage or injury resulted more overtly from the illegal or unethical actions of a third party (such as abusive users on the platform, foreign operatives set on electoral interference, or third-party advertisers illicitly collecting or sharing users' personal data), rather than directly from the platform itself.[264] Moreover, "Courts are reluctant to hold a defendant liable when the loss is triggered by the deliberate and often criminal act of a third person over whom the defendant has no control."[265] This suggests that even provided a particular platform-facilitated systemic harm is shown to meet all the other elements of negligence in tort, the platform company may be found not liable due to the third-party conduct that ultimately caused the harm in question, especially if the conduct was intentional and malicious.

However, courts have also established that a "wrongful act by the third party will not break the chain of causation where the breach of the [defendant's] obligation initiated the chain of events leading to the loss [or injury] and the breaching party must account for this loss [or injury] in full, subject to any issue of apportionment."[266] On this basis, a marginalized group that has systemically sustained disproportionate or targeted harm from third-party actors on a platform might argue that the platform company is responsible for having "initiated the chain of events" through the components of emergent systemic harm discussed above creating an environment that fostered such third-party conduct, without taking any measures to

[261] *Ibid* at para 47.
[262] Zeynep Tufekci, "Why Zuckerberg's 14-Year Apology Tour Hasn't Fixed Facebook", *Wired* (6 April 2018), online: <https://www.wired.com/story/why-zuckerberg-15-year-apology-tour-hasnt-fixed-facebook/>.
[263] "It is true that a person who commits a fault is not liable for the consequences of a new event that the person had nothing to do with and that has no relationship to the initial fault. This is sometimes referred to as the principle of *novus actus interveniens*: that new event may break the direct relationship required under art. 1607 C.C.Q. between the fault and the injury. Two conditions must be met for this principle to apply, however. First, the causal link between the fault and the injury must be completely broken. Second, there must be a causal link between that new event and the injury. Otherwise, the initial fault is one of the faults that caused the injury, in which case an issue of apportionment of liability may arise." *Salomon v Matte-Thompson*, 2019 SCC 14, at para 91.
[264] In fact, in some sense it may be possible to view the principle of intermediary liability and section 230 of the *Communications Decency Act* as pre-emptive enshrining of *novus actus interveniens* with respect to online platforms.
[265] Osborne, *supra* note 220 at 108.
[266] *Cole Parliament et al v DW Conley and V Park*, 2019 ONSC 3996, at para 27.

prevent or mitigate the risk of the foreseeable harm that occurred. A specific example of this might a group of racialized LGBTQ+ content creators bringing action for systemic negligence by YouTube enabling intentional infliction of mental distress. This claim would be based on YouTube's recommendations algorithm, ad-based business model, incentives for creators, and known user base turning the platform into a system with foreseeably harmful operations and practices, resulting in such creators being targeted for sustained, public race- and sexuality-based harassment, which in one specific case YouTube determined it was unable to address, before reconsidering in light of public pressure.[267]

At this point, some clarification may be necessary with regard to the distinction between emergent systemic harms and what this paper earlier defined as non-emergent systemic harms, particularly given that many of the examples above focused on situations that involve abusive users targeting members of marginalized groups (which would be a *non-emergent* systemic harm due to the involvement of a "non-innocent" component, the abusive users). The distinction is, to some extent, a question of scale. One platform user being abused or harmed by another individual platform user would not provide grounds to bring a negligence action against the platform, and would also be barred under principles enacted in section 230 of the *Communications Decency Act* or similar laws in other jurisdictions. However, dozens or hundreds of people belonging to a particular group, being harmed in a particular way, would constitute a systemic issue and thus could ground a claim for systemic harm as a result of platform negligence. The platform is not being held liable for one user abusing or harming another user, but rather, for building, maintaining, and upholding (and profiting from) platform-wide conditions, affordances, incentives, and system interactions that fostered and sustained an emergent—or non-emergent, depending on point of view—systemic harm to a particular marginalized class of individuals. The underlying theory is that of systemic negligence (responsibility for the system and the outcomes it is tilted towards, on a structural level), rather than intermediary or vicarious liability (responsibility for a specific wrongdoer or wrongdoing in any single given case).

C. An Intersectional and Relational Recalibration of Reasonable Foreseeability

While the analysis in the previous section attempted to demonstrate how traditional foreseeability considerations might operate in the context of negligence liability for platform-facilitated emergent systemic harms, this present section argues for a more intersectional, comprehensive, and relational understanding of reasonable foreseeability. The objective is to develop a more contemporary understanding of what harms and risks should be considered reasonably foreseeable in today's society, with the normative and binding force of tort law, in a way that incorporates principles of equality, intersectionality, power imbalances, historical patterns of oppression, and systemic discrimination.

i. Intersectional Foreseeability and the Reasonable Person

An updated understanding of foreseeability should correspondingly influence other elements of negligence that rely on foreseeability, such whether or not a particular duty of care exists and what constitutes a reasonable standard of care. As Peppin states,

> Elements of negligence, such as recognition of harm, duty, and the standard of care, need to be construed in a manner sensitive to historic disadvantage. Hierarchy, dominance, and disadvantage

[267] Kevin Roose, "A Thorn in YouTube's Side Digs In Even Deeper", *New York Times* (12 February 2020), online: <https://www.nytimes.com/2020/02/12/technology/carlos-maza-youtube-vox.html>.

> characteristic [*sic*] relationships in life; awareness of this dimension should enter into the determination of liability and compensation.[268]

With respect to platform-facilitated emergent systemic harms to marginalized groups, a more comprehensive and inclusive concept of reasonable foreseeability, as applied to the types of platform-facilitated systemic harms and risks disproportionately visited upon marginalized communities, is one way the law can respond to the platform foreseeability gap and encourage more equitable developments in the area of digital platform policy and practice. This might be considered *intersectional foreseeability*.[269]

Applying this kind of critical intersectional lens to negligence and tort law more broadly is not a novel exercise, but rather, follows in the footsteps of a pre-existing body of literature in feminist tort law scholarship, as well as critical race and critical Indigenous analyses of Canadian tort law concepts. Leslie Bender's 1993 article, "An Overview of Feminist Torts Scholarship"*,* revisits and extends her earlier work advancing the argument that

> [F]eminist theory encourages us to challenge traditional modes of legal analysis and to rethink the questions we ask, including: who are the parties involved, whose interests are protected by tort law, what are appropriate forms of compensation, how should we allocate responsibility for harms and risks, and what assumptions and values underlie various tort doctrines? […] Feminist legal theories can influence the kinds of practices that are permissible for large, profit-seeking corporate enterprises[.][270]

Sherilyn J. Pickering provides an overview of various feminist legal theories as applied to key doctrines in tort law. The schools of thought she canvasses highlight, interrogate, and debate with each other over, for instance, what is presented as the gendered nature of tort law and its stated priorities or framing of values;[271] the inequitable allocation of burden of proof (for instance, where the defendant is a well-heeled corporation);[272] what should form the basis and substance of a duty of care that assumes and supports a more caring society;[273] the appropriate standard of care in a society that recognizes interconnectedness between individuals while not unduly impinging on individual autonomy;[274] and how to achieve a more

[268] Pickering, *supra* note 248 at 238-39, citing Patricia Peppin, "A Feminist Challenge to Tort Law" in Anne Bottomley, ed, *Feminist Perspectives on the Foundational Subjects of Law* (London: Cavendish Publishing Limited, 1996) 69.

[269] "'Intersectionality' was coined in 1989 by Kimberlé Crenshaw, a civil rights activist and legal scholar. In a paper for the University of Chicago Legal Forum, Crenshaw wrote that traditional feminist ideas and antiracist policies exclude black women because they face overlapping discrimination unique to them. 'Because the intersectional experience is greater than the sum of racism and sexism, any analysis that does not take intersectionality into account cannot sufficiently address the particular manner in which Black women are subordinated'". Merill Perlman, "The origin of the term 'intersectionality'" (23 October 2018), online: *Columbia Journalism Review* <https://www.cjr.org/language_corner/intersectionality.php>, citing Kimberle Crenshaw, "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics" (1989) 1:8 University of Chicago Legal Form 139. The term "intersectionality" and derivatives (e.g. "intersectional") has since been extended in common usage to apply to situations where other marginalized identifies overlap within an individual, in addition to being both Black and a woman.

[270] Leslie Bender, "An Overview of Feminist Torts Scholarship" (1993) 78 Cornell L Rev 575 at 581.

[271] See e.g., Pickering, *supra* note 248 at 236; see also *Ibid* at 578.

[272] "If all cases were analyzed within their appropriate context, it would enable the judiciary to place the burden of proof of causation on the appropriate party. Bender writes that in instances where the defendant is a corporation, it has a far greater ability to bear the burden of proof than the plaintiff." Pickering, *supra* note 248 at 238 (footnote omitted)

[273] See e.g., Pickering, *supra* note 248 at 241

[274] See e.g., Pickering, *supra* note 248 at 246.

feminist version of tort law without falling back on outdated and harmful gender stereotypes and gender essentialism.[275]

Similarly, in the context of interpreting tort law to take into account systemic racism and structural power imbalances, Ruparelia refers to Mari Matsuda's position regarding causation:

> [W]hen determining legal cause, tort law must adopt a more expansive view. In [Matsuda's] view, current rules employ limiting doctrinal devices that are artificial and reflect policy decisions that protect selective interests, largely economic ones. Instead, she suggests that we consider causation from the perspective of defendants' ability to avoid, prevent and remedy social harm, such that persons in the best position to prevent the harm are held accountable.[276]

For the purposes of recalibrating foreseeability in the context of platform-facilitated emergent systemic harms, this section will rely in particular on critiques and re-imaginings of "the reasonable person". The "reasonable person" is a core concept in negligence law that undergirds the same three elements of analysis that turn on reasonable foreseeability: assessing whether a duty of care exists;[277] determining whether a particular type of harm was foreseeable;[278] and evaluating what conduct is required to meet the standard of care in a given situation.[279] In short, what is considered "reasonably foreseeable" in law is what the law believes the much invoked "reasonable person" would foresee. Whom precisely the courts think this "reasonable person" is— including this person's imagined values, perspectives, life experiences, knowledge base, and ethical sensibilities—thus wields much power to shape the law and societal norms around negligence,[280] harms, risk, and what constitutes (ir)responsible conduct in interpersonal, commercial, and sociopolitical contexts.

Some critiques of the reasonable person in law parallel critiques of Silicon Valley executives becoming arbiters of sociopolitical mores through their platforms, including the observation that the reasonable person is still to a large extent rooted in its original conception of "the reasonable man", a standard "created and originally argued, theorized, and written about"[281] by men (moreover, a specific privileged subset of men with particular socioeconomic standing and power). In the context of platform-facilitated emergent systemic harm, this critique of the reasonable man/person brings to mind similar criticisms of the homogeneity of Silicon Valley, particularly among the founders, executives, engineers, and managers

[275] See e.g., Pickering, *supra* note 248 at 243-44, 245.
[276] Ruparelia, *supra* note 221 at at 101.
[277] "A duty of care only arises where the circumstances of time, place and person would create in the mind of a reasonable person in those circumstances such a probability of harm resulting to other persons as to require him or her to take care to avert that probable result." *Garratt v Orillia Power Distribution Corp*, 2008 ONCA 422, at para 46 (in-line citations omitted) [*Garratt*].
[278] "The [foreseeability] inquiry being objective (that is, into what reasonably *ought* to have been foreseen), it must be undertaken from the standpoint of a reasonable person. Whether, therefore, the defendant actually foresaw the risk which ultimately manifested in injury to the plaintiff is not determinative." *Rankin's Garage*, *supra* note 237 at para 77 (footnotes omitted, emphasis in original).
[279] "The standard of care must be appropriate in the circumstances, and is that of the ordinary, reasonable, cautious and prudent person in the position of the defendant. ... Conduct is negligent if it creates an objectively unreasonable risk of harm [Ryan v. Victoria (City)]. The factors to be considered by a reasonable person when deciding upon a course of conduct include: the likelihood that its conduct will injure others, the seriousness of the injury if it happens, balanced against the interest which must be sacrificed to avoid the risk." *Garratt*, *supra* note 277 (citing trial judge, at para 21).
[280] "The reasonable person is critical to negligence liability for he determines fault and thus sets an important threshold for liability." Mayo Moran, *Rethinking the Reasonable Person: An Egalitarian Reconstruction of the Objective Standard* (New York: University of Oxford Press, 2003) at 14, cited in Pickering, *supra* note 248 at
[281] Pickering, *supra* note 248 at 239, citing Leslie Bender, "A Lawyer's Primer of Feminist Theory and Tort" in D. Kelly Weisberg, ed., *Feminist Legal Theory Foundations* (Philadelphia: Temple University Press, 1993) 58 at 67.

whose values infuse the digital platforms that govern the rest of us.[282] In particular, the following statement by Moran, regarding the traditionally unrepresentative nature of "the reasonable person", highlights the same issue as that which gives rise to the platform foreseeability gap: "behaviour that may well be reasonable from some *other* perspective may fail to be identified as reasonable because of the narrowness of the perspective enshrined in the law."[283] The same could be said for reasonable foreseeability: risks and harms that may be reasonably foreseeable from "some *other* perspective" (as opposed to perspectives of those who are not considered "other") may fail to be identified as such because of the narrowness of the perspective enshrined in how digital platforms and their business models and ecosystems are developed, designed, built, managed, and marketed. Peppin's warning for tort law applies to the context of platform-facilitated systemic harms, as well: "[A]s long as this legal standard continues to reflect the dominant view, it will continue to require conformity and to exclude alternative experiences. It will continue to reflect the 'history of bias in favour of white male, middle-class, ablebodied values…'"[284]

In response to such concerns, Naomi Cahn advocates for "a context-based standard of care that would enable minorities to voice their perspectives and experiences."[285] Similarly, Ruparelia proposes that "the standard of care should be that of the reasonable non-racist (non-sexist, non-homophobic, nonableist) person in similar circumstances."[286] She elaborates:

> As a starting point, an anti-racist framework needs to be incorporated into existing negligence actions. This should be inserted at every stage of the negligence analysis. However, and by way of illustration, it is easiest to envision this incorporation at the standard of care determination. When we consider what a reasonable person would have done [or foreseen] in particular circumstances, we must begin from the perspective that reasonable people are not racist and that racist acts are neither reasonable nor socially useful. This should be true in any claim of negligence, not only those in which discrimination is specifically asserted.[287]

More than a presumption that the reasonable person is not racist, however, the platform foreseeability gap reveals that it is just as important to incorporate, as part of the analytical framework for negligence, a presumption that the reasonable person understands *being vulnerable to those who* are *racist* (as well as those who are ableist, sexist, transphobic, and so forth).

[282] See e.g., "Today, of course, hateful, enraging words are routinely foisted on the public by users of all three companies' products, whether in individual tweets and Facebook posts or in flawed Google News algorithms. Championing freedom of speech has become a business model in itself, a cover for maximizing engagement and attracting ad revenue, with the social damage mostly pushed aside for others to bear. When the Internet was young, the reason to clean it up was basic human empathy—the idea that one's friends and neighbors, at home or on the other side of the world, were worth respecting. In 2017, the reason is self-preservation: American democracy is struggling to withstand the rampant, profit-based manipulation of the public's emotions and hatreds." Noam Cohen, "The Origin of Silicon Valley's Dysfunctional Attitude Toward Hate Speech", *New Yorker* (28 November 2017), online: < https://www.newyorker.com/tech/annals-of-technology/origin-silicon-valley-dysfunctional-attitude-toward-hate-speech>. For a more legal perspective of the generally homogenous view towards this issue among major U.S.-based platform companies, see Klonick, "The New Governors", *supra* note 5.

[283] Pickering, *supra* note 248 at 239, citing Moran, *supra* note 280 (emphasis in original).

[284] Pickering, *supra* note 248 at 239, citing Peppin, *supra* note 268.

[285] Pickering, *supra* note 248 at 243 (footnote omitted)

[286] She elaborates, "There are many ways that the courts could develop this standard of non-discrimination. These include by referring to the *Charter*, particularly section 15, as well as to human rights codes, reports, guidelines, policies, scholarly research and jurisprudence. Although a breach of human rights statutes cannot be used to create a statutory tort, a breach of those codes could be viewed as evidence of breach of the standard of care for the purposes of negligence. As well, *Charter* values are understood to inform common law principles." Ruparelia, *supra* note 221 at 100 (footnotes omitted).

[287] Ruparelia, *supra* note 221 at 97 (footnotes omitted, emphasis added).

Nowhere does the distance between the seemingly invulnerable reasonable person and the all too vulnerable members of marginalized communities strike more forcefully than in Hadley Friedland's exposition of the reasonable Cree person:

> The reasonable Cree person is just this—an ordinary legal actor who uses the Cree legal tradition as a practical guide to think through and make reasonable and principled decisions, when, like the rest of us, she is called to judgment. […] The reasonable Cree person's thinking reflects a long history of Cree legal thought and experience, but also the current political, social, economic and natural realities of today and legal relationships with other peoples. […]
>
> I can identify nothing about our Cree reasonable person that would protect her from suffering the fate of far too many Indigenous women and girls across Canada. Nothing.[288]

As made painfully clear by Friedland, the truly representative reasonable person would not be one who is vaguely aware of, but ultimately above, the realities of living in the trenches of a deeply and violently unequal society. The reasonable person has held a shovel. The reasonable person certainly should not be implicitly modelled after those who have been historically and most frequently responsible for targeting, launching, and sustaining the artillery blasts.

As expressed in the tweet quoted in the opening to this paper,[289] that marginalized individuals appear to possess greater foresight and be more perceptive of how things can go wrong is not a magical superpower, but more often reflects the rather sadder reality of hard-won perspective and vigilance developed out of firsthand experience and necessity. Given the difficulties of an abstract legal fiction experiencing the material reality of discrimination firsthand, a more feasible substitute might involve attributing to the reasonable person, and incorporating into the legal standard of what is reasonably foreseeable, perspicacious understanding of what it means to be, and the different ways someone can be, vulnerable to systemic discrimination and hate-based abuse. In fact, such an understanding of the reasonable person has already been articulated and enforced by the Supreme Court of Canada:

> The reasonable person must be taken to be aware of the history of discrimination faced by disadvantaged groups in Canadian society protected by the *Charter's* equality provisions. These are matters of which judicial notice may be taken. […]
>
> The reasonable person is not only a member of the Canadian community, but also, more specifically, is a member of the local communities in which the case at issue arose (in this case, the Nova Scotia and Halifax communities). Such a person must be taken to possess knowledge of the local population and its racial dynamics, including the existence in the community, of a history of widespread and systemic discrimination against black and aboriginal people, and high profile clashes between the police and the visible minority population over policing issues […]
>
> We conclude that the reasonable person contemplated by de Grandpré J., and endorsed by Canadian courts is a person who approaches the question of whether there exists a reasonable apprehension of bias with a complex and contextualized understanding of the issues in the case. … The reasonable person is cognizant of the racial dynamics in the local community, and, as a member of the Canadian community, is supportive of the principles of equality.[290]

Based on the above academic literature and jurisprudence, it seems that if this reasonable person described by the Court were in charge of designing and building digital platforms with the power and reach of today's largest technology companies, the platform foreseeability gap would be much smaller than it has proven to be over the years, or even non-existent. It is this intersectionality-informed

[288] Hadley Friedland, "Navigating through Narratives of Despaire: Making Space for the Cree Reasonable Person in the Canadian Justice System" (2016) 67 UNBLJ 269 at 275-76, 280, 281.

[289] @CaseyExplosion, *supra* note 2.

[290] *R v S (RD)*, [1997] 3 SCR 484, 118 CCC (3d) 353, at paras 46-48.

reasonable person who should form the basis of determining the reasonable foreseeability of certain kinds of harms and risks to members of marginalized communities, as well as the appropriate standard of care digital platform companies should meet so as to not be considered legally negligent.

ii. Relational Foreseeability

Applying a standard of foreseeability that is explicitly intersectional, when engaging in a negligence analysis, involves understanding foreseeability as *relational.* In some ways, it is redundant to speak of "relational foreseeability", or any aspect of tort law being particularly relational, because all of tort law is an inherently relational area of law. The entire purpose of tort law is to govern and manage relationships between private individuals, organizations, and corporate actors in society, including providing mechanisms to remedy or repair relationships where one party has broken or disrupted a relationship by causing injury to the other. In this vein, tort law is founded on recognizing that people and entities must have principles by which to navigate the inevitable mishaps and hazards that arise from living in proximity to one another, whether that is the physical proximity of living beside each other, or the conceptual proximity of online transactions and interactions through digital platforms. However, it is still worth discussing relational foreseeability as a standalone concept, particularly with respect to platform-facilitated emergent systemic harms and potential application of negligence liability.

What characterizes relational foreseeability, I propose, is its critical focus on situating tort victims, plaintiffs, defendants, and any relevant third parties within a web of inter-constituting relationships that takes into account power differentials and how relationships are structured, including the particular impacts on different groups of people within those webs. This concept of relational foreseeability builds upon and is inspired and informed by Carys Craig's work on "relational copyright" and Ian Kerr's work on "relational privacy" (which was in turn also inspired by Craig). Both relational copyright and relational privacy emerged from applying a critical feminist lens to traditional copyright and traditional privacy doctrines and theories, respectively, similarly to some of the work done by feminist tort theorists as discussed above.

Relational copyright encompasses understanding the following, as developed by Craig:

> [Copyright] structures relationships between authors and users, allocating powers and responsibilities amongst members of cultural communities, and establishing the rules of communication and exchange. To assess the nature of copyright with reference only to the copyright owner's private sphere of entitlement is to undermine its normative significance. The importance of copyright lies in its capacity to structure relations of communication, and also to establish the power dynamics that will shape these relations. Its purpose is to maximize communication and exchange by putting in place incentives for the creation and dissemination of intellectual works. Relational feminism can teach us that an individualized account of the copyright holder's right will disregard the significance of the relationships affected by copyright and will be blind to the power dimensions and social implications of the copyright system."[291]

Central to relational copyright is the rejection of the traditional romanticized notion of the author as embodying "an individual's genius and greatness particular to him"[292]—with implicit assumptions of isolated individualism, mystical detachment from or transcendence above day-to-day society and politics, and concealment of power structures that the copyright holder benefits from and can help to perpetuate—in favour of recognizing that creators are one node in a web of culture, creation, and communications, who inherently operate and create while in constant dialogue with other nodes (other people and institutions). The creator or copyright holder, and their work, cannot thus be separated from other nodes

[291] Carys J Craig, "Reconstructing the Author-Self: Some Feminist Lessons for Copyright Law" (2007) 15:2 J Gender Soc Pol'y & L 207 at 263.
[292] *Ibid* at 213.

or treated as if they are not themselves embedded in society's web and implicated within the network of sociopolitical relations, their dynamics, and the ways in which such relations are structured and governed, or in turn structure and govern other nodes, including the creators or copyright holders themselves.

Similarly, according to Kerr, relational privacy is a term that emphasizes the role that privacy plays in structuring relationships between people, as well as how relational considerations should inform legal and other analyses of privacy, including in the context of robotics and artificial intelligence:

> Although the moniker of relational privacy has been used to mean different things, I use it here to refer broadly to any account of privacy that coincides, borrows, or builds upon broader relational theory in the tradition of Jennifer Nedelsky and other feminist thinkers who "insist on the centrality of relationships in human lives" and recognize that privacy values "require structures of relationships that support them — that allow people the opportunities to retreat from others in various ways." On this view, "people are not self-made." Instead, they are constituted by networks of nested relationships. Their autonomy "can thrive or whither . . . depending on the structures of relationships they are embedded in."[293]

Just as with relational copyright, the core concept of relational privacy is the focus on people being nested within networks of relationships in which they constitute and are constituted by one another, with the corollary point that being "self-made" is a myth.

To the above two concepts I add a third taken from Frank Pasquale: relational rule of law (a term not from him but which I have somewhat taken the liberty to apply). According to Pasquale, the automation of legal services through artificial intelligence, much enthused by "legal futurists" and the "legal tech" industry, would be a "betrayal of the rule of law" rather than efficient and unproblematic enforcing of rule of law.[294] To support this contention, Pasquale draws on a more expansive "Legal Process conception" of the rule of law, in opposition to legal tech's "incomplete normative account of the rule of law" which is a narrower historicist or formalist conception of rule of law.[295] Notably for the purposes of developing relational foreseeability, and read next to relational copyright and relational privacy, the "Legal Process conception" of rule of law involves the following:

> While [historicist and formalist] approaches emphasise the "rule" side of the rule of law, the Legal Process approach emphasizes "law" as its core component. Law *as a social institution is multi-faceted and embedded in particular political systems and traditions*, such as rights to appeal and explanations for decisions. […]
>
> When conflicts over interpretation arise, the Legal Process approach to the rule of law demands the clashing parties are offered "reasoned elaboration of *the connection between recognized, pre-existing sources of legal authority and the determination of rights and responsibilities in particular*

[293] Kerr, "Schrödinger's Robot", *supra* note 12 at 130 (footnotes omitted).

[294] Frank Pasquale, "A Rule of Persons, Not Machines: The Limits of Legal Automation" (2019) 87:1 Geo Wash L Rev at 59 [Pasquale, "Rule of Persons"].

[295] "The legal futurists' partial vision of economic progress reflects a similarly incomplete normative account of the rule of law—one that asks both too much, and too little, of legal institutions. Whatever other normative goals judges and regulators pursue, they should adhere to the rule of law. Richard Fallon has observed that there are at least three distinct ideal-typical accounts of the rule of law in contemporary jurisprudence. Legal automators tend to focus on historicist accounts (which associate the rule of law with "rule by norms laid down by legitimate lawmaking authorities prior to their application to particular cases") and formalism (which defines 'the ideal if not necessary form of 'law'' as "that of a 'rule,' conceived as a clear prescription that exists prior to its application and that determines appropriate conduct or legal outcomes.'). Were federal health privacy regulation really reducible to 'requirements extraction' encoded in software, that encoding would amount to a real advance for the rule of law, in its historicist and formalist conceptions. The law would be as executable as a software command. Similarly, the translation of traffic rules into a series of chatbot prompts renders the law into a crystalline form—if not application." *Ibid* at 48 (footnotes omitted).

> *cases*"—not simple disposition of their cases via code. […] One-sided deployments of vastly superior legal-technological resources also undermine *the types of dialogue and fair play* valued by the Legal Process school.[296]

Pasquale concludes his article by emphasizing, "The rule of law *entails a system of social relationships* and legitimate governance, not simply the transfer and evaluation of information about behavior."[297] In other words, rule of law must also be considered through a relational lens in order to achieve a more robust understanding of its function and mechanisms, which can ground more fulsome and accurate analyses, policies, and practices that purport to turn on the concept. Just like copyright and privacy, rule of law does not exist in a sociopolitical vacuum.

Neither does foreseeability. Drawing on the relational concepts of law above, relational foreseeability demands that the reasonable foreseeability analysis in negligence recognize that those in defendant positions are also nodes within broader networks of relationships structured in particular ways, both influenced by and drivers of sociopolitical and cultural factors, and with two-way, yet often one-sided impacts on other, more vulnerable nodes. While this premise may seem obvious in more typical negligence cases (such as personal injury, a tort lawsuit between two private individuals, or a product liability case against a traditional manufacturer), the relational aspect of foreseeability—and negligence generally—warrants particular emphasis in the context of digital platform-facilitated emergent systemic harms specifically. This is because the "self-made solo genius artist" that Craig describes at the heart of traditional copyright doctrine bears a striking resemblance to the "boy kings",[298] "loner genius nerds",[299] and "brilliant jerks"[300] of Silicon Valley, enshrined alongside the myth of the god-like tech founder.[301] These popular and continually reified tropes feed directly into the technology sector's diversity and equity problems,[302] which in turn unjustifiably and unnecessarily exacerbate conditions that foster the platform foreseeability gap.[303] "The habit of treating artists [and technology company founders, engineers,

[296] *Ibid* at 49-50 (emphasis added).

[297] *Ibid* at 60 (emphasis added).

[298] Katherine Losse, *The Boy Kings: A Journey Into The Heart Of The Social Network* (New York: Free Press, 2012).

[299] Claire Cain Miller, "Tech's Damaging Myth of the Loner Genius Nerd", *New York Times* (12 August 2017), online: <https://www.nytimes.com/2017/08/12/upshot/techs-damaging-myth-of-the-loner-genius-nerd.html>.

[300] Shira Ovide, "What If Silicon Valley's 'Brilliant Jerks' Are Just Jerks?" *Washington Post* (30 August 2019), online: <https://www.washingtonpost.com/business/what-if-silicon-valleys-brilliant-jerks-are-just-jerks/2019/08/29/c605b8f0-ca70-11e9-9615-8f1a32962e04_story.html>.

[301] See e.g., Hamish McKenzie, "There's No Such Thing as Individual Genius in Silicon Valley", *Pando* (10 April 2012), online: <https://pando.com/2012/04/10/theres-no-such-thing-as-individual-genius-in-silicon-valley/>; University of Virginia, "Debunking the Tech Founder Myth" (16 September 2019), online: *Communications of the ACM* <https://cacm.acm.org/careers/239476-debunking-the-tech-founder-myth/fulltext>; Ellen Pao, "Tech Founders' Absolute Power Is Destroying Company Culture", *Wired* (10 October 2018), online: <https://www.wired.com/story/ellen-pao-founders-absolute-power-destroying-company-culture/>; and "Shanley Kane, "The Startup Mythologies Trifecta", *Model View Culture* (17 March 2014), online: <https://modelviewculture.com/pieces/the-startup-mythologies-trifecta>.

[302] "These mythologies all work in concert. The tropes elevating programming talent feed into the prestige given to founders. The glamorous image of founders and startups helps feed the myth around the 'get rich quick' trap of startup equity. The privileging of technical talent influences who gets to become founders; and those founders help determine who even has a shot at what is often termed the 'startup lotto.' ... Ultimately, understanding what the mythologies are vs. the reality, how these myths are created and propagated, how they work together, and how they disproportionately punish and exclude minorities in tech is critical to tearing down some of the structure underlying tech's inequalities." Kane, *supra* note 301.

[303] "A cult of genius can be toxic. It excuses bad behavior and allows prejudices to be cloaked in subjective assessments of intelligence and value. Today it is at the heart of the tech world's problematic culture, and linked to many key issues: tech bros' white male homogeneity, rampant sexual harassment, and focus on catering to the

managers, and executives] as transcendent creators rather than as players in an economic system serves to protect them from typical workplace expectations."[304] Relational foreseeability recognizes that platform companies and their owners, managers, and employees (both star and low-level) are indeed players on the ground in various economic, political and other systems, and that they are enmeshed in a sociopolitical web of inter-constituting relationships along with the rest of us. Alongside workplace expectations, tort law should not protect them from well-established legal expectations with respect to possessing reasonable foresight of potential risks and harms to marginalized and vulnerable users of their products, and meeting the appropriate standard of care to prevent or mitigate those harms.

Part V. Foreseeability in the Time of Emergence

This paper began with the premise of "emergent systemic harms" arising from digital platform ecosystems, based on the rhetoric of "unintended consequences", and subsequently demonstrated that many such harms were not as "emergent" as they seemed at first blush. This destabilization of emergent systemic harm and the ostensibly unintended nature of platform-facilitated systemic harms informed a re-evaluation of the reasonable foreseeability test in negligence law, which advocated the application of intersectional and relational foreseeability, as well as an equality-and-discrimination-informed reasonable person standard, to account for and encourage closure of the "platform foreseeability gap" by those with decision-making power in digital platform companies. For the sake of completion, Part V will analyze the circumstance where a platform-facilitated systemic harm was in fact truly emergent in all senses of the term and genuinely unforeseeable. Such a situation either already exists or is increasingly likely, due to continuous adoption of and advancements in artificial intelligence (AI) and machine learning algorithms by digital platform companies to fulfil various functions in running their platforms.

Much has already been written on the problems that AI, autonomous algorithms, and unsupervised machine learning will pose to the traditional legal test used to determine negligence-based liability, largely based on the presumed loss of foreseeability. Difficulties include the unintelligibility of AI-driven technologies to even their own designers and developers;[305] the potential lacuna of remedy for victims of unexpected algorithm-induced harms in the absence of a new or updated legal mechanism to provide for accountability and redress;[306] the "collapse" of control in addition to foreseeability, with respect to increasingly autonomous algorithms;[307] and the black box nature and complexity, incomprehensible to humans, of how algorithms process, draw inferences from, and make decisions based on ingested data, again impairing or precluding foreseeability.[308] As Bathaee sums up,

> In the case of black-box AI, the result of the AI's decision or conduct may not have been in any way foreseeable by the AI's creator or user. For example, the AI may reach a counter-intuitive solution, find an obscure pattern hidden deep in petabytes of data, engage in conduct in which a human being could not have engaged (e.g., at faster speeds), or make decisions based on higher-dimensional relationships between variables that no human can visualize. Put simply, if even the creator of the AI cannot foresee its effects, a reasonable person cannot either. Indeed, if the creator of AI cannot

concerns of the most privileged in society." Olivia Goldhill, "Philosophers are the original tech bros" (16 April 2017), online: *Quartz* <https://qz.com/959409/philosophers-are-the-original-tech-bros/>.

[304] Amanda Hess, "How the Myth of the Artistic Genius Excuses the Abuse of Women", *New York Times* (10 November 2017), online: <https://www.nytimes.com/2017/11/10/arts/sexual-harassment-art-hollywood.html>.

[305] Sullivan & Schweikart, *supra* note 19 at 160-61.

[306] *Ibid* at 163.

[307] *Ibid* at 163.

[308] Bathaee, *supra* note 19 at 891-92.

> necessarily foresee how the AI will make decisions, what conduct it will engage in, or the nature of the patterns it will find in data, what can be said about the reasonable person in such a situation?[309]

In response to what appears to be an inevitable AI foreseeability (and causation) crisis, legal scholars and policy experts have explored a range of potential solutions for determining and allocating liability for harm, based on various pre-existing legal doctrines. These include, broadly speaking: applying strict liability (where demonstration of damage alone is sufficient to ground liability, regardless of intent or foreseeability);[310] control theories such as principal-agent liability or vicarious liability (where the algorithm is the "agent" or "employee" of the business or programmer that deployed it);[311] no-fault insurance pools funded by the AI industry as a whole, possibly on the basis of common enterprise liability;[312] *res ipsa loquitur* ("the thing speaks for itself", where the fact alone that harm occurred is taken as sufficient evidence that something was awry and the defendant should be liable for it[313]—this strategy is not an option in Canada, as the Supreme Court of Canada put the *res ipsa loquitur* doctrine to rest in *Fontaine v British Columbia Office Administrator*[314]); and a suite of sliding scale options based in traditional tort law, which begin with traditional intent and negligence-based liability for less autonomous algorithms operated with the most transparency, up to modified negligence and then strict liability for the most autonomous algorithms operated with the least transparency.[315]

A. Sliding Scale of Tort Liability Standards

As a tentative preliminary view, applying a sliding scale of liability standards as Bathaee suggests would seem to be the most suitable approach to platform-facilitated emergent systemic harm to marginalized groups, where unsupervised autonomous algorithms form part of the system. A one-size-fits all legal standard may be unworkable or lead to poor outcomes in many cases, given the vast range of purposes for which platform algorithms may be used and the variety of contexts and environments in which the algorithms would operate—including by and within a single platform, let alone among many different ones. Bathaee applies the sliding scale framework to the entire spectrum of algorithms, from supervised non-autonomous algorithms up to unsupervised autonomous algorithms. However, the flexibility and context-sensitivity of such an approach would seem to offer value even upon narrowing the problem set to exclusively unsupervised autonomous algorithms, the kind that would be at the root of truly emergent and unforeseeable platform-facilitated systemic harm.

Without purporting to set out a formal test, three proposed standards in particular stand out as suitable candidates for the scale of liability as applied to unsupervised black-box platform algorithms,[316] in ascending order of liability: a no-fault insurance pool funded by the AI industry,[317] a modified form of the

[309] *Ibid* at 924 (footnotes omitted).

[310] See e.g., Jennifer A Chandler, "Negligence Liability for Breaches of Data Security" (2008) 23 Banking & Finance Law Review 223 at 230; Bathaee, *supra* note 19 at 932; Jackson, *supra* note 19 at 60; and David C Vladeck, "Machines Without Principals: Liability Rules and Artificial Intelligence" (2014) 89:1 Wash L Rev 117 at 146.

[311] See e.g., Bathaee, *supra* note 19 at 935-36, Jackson, *supra* note 19 at 55-56; and Vladeck, *supra* note 310 at 122.

[312] See e.g., Gary Lea, "Who's to blame when artificial intelligence systems go wrong?", *The Conversation* (16 August 2015), online: <https://theconversation.com/whos-to-blame-when-artificial-intelligence-systems-go-wrong-45771>; Sullivan & Schweikart, *supra* note 19 at 164, Vladeck, *supra* note 310 at 129 and 149.

[313] See e.g., Jackson, *supra* note 19 at 59,

[314] [1998] 1 SCR 424.

[315] See e.g., Sullivan & Schweikart, *supra* note 19 at 164; Bathaee, *supra* note 19 at 932-38; and Andrea Bertolini, "Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules" (2013) 5:2 Law, Innov & Tech 214 at 234.

[316] Bathaee, *supra* note 19 at (diagram)

[317] See e.g., Lea, *supra* note 312; and "Here, of course, there would be no 'wrongdoers.' There would instead be an inference of liability drawn by operation of law to protect a blameless party (the person who sustained injured) by

traditional negligence test, and strict liability. To determine which liability standard to apply, a court would have to first determine the level of risk and harm that was at stake in the context of the algorithm's function and operation: low stakes, medium stakes, or high stakes.[318] This decision would take into account both the type of harm that in fact occurred, and the level of risk that would have been apparent at the start, as a result of the platform's purpose and the algorithm's role in it.

However, making such a determination poses particular difficulties in the context of emergent systemic harm, where the very premise is that individual components of the platform system (including the algorithm) to all appearances are low-risk, until they collectively prove otherwise. Some platforms may be identifiable as involving medium or high risk from the start and then leading to high-stakes actual harm, such as a (hypothetical) healthcare platform that purports to diagnose various maladies using machine learning algorithms, based on user-submitted photos and videos, and which leads to widespread misdiagnoses that systematically occur with particular types of infections most common to women. Other platforms, though, may more resemble the ones examined in this paper, where an initially low-risk-seeming venture leads to either, or both, low-stakes and high-stakes systemic harm. One imagined example of this is an online platform developed exclusively for creating and sharing jokes and various forms of comedy, with crowdsourced and user-generated content both used as input data for an algorithm trained to "learn" humour, and the algorithm's output attempts at humour are provided back to the community for users to incorporate into their own material. In one scenario, those on the platform are eventually fooled by a platform-wide hoax unexpectedly generated by the algorithm as a "prank", resulting in moderate monetary losses among all users—a low to medium-stakes systemic harm. In another scenario, that same platform works the same way, but over time researchers discover a form of psychological trauma among the platform's users who belong to historically marginalized communities, and establish that it is more likely than not connected to being active and engaged users on that platform. In this case, a seemingly low-risk platform resulted in high-stakes systemic harm.

Traditional legal principles would suggest that it would seem unjust and arbitrary to impose a standard of liability based on consequence alone, regardless of the defendant's conduct and when both consequences were unforeseeable at the start. It is possible, however, that the introduction of (unsupervised and autonomous) algorithms necessitates as significant a shift in such legal principles, as these technological advances engendered or will cause in so many other spheres of society itself. That the comedy platform may rely on the no-fault insurance pool in the first scenario, but must face negligence or strict liability in the second scenario, represents the law shifting not based on the defendant, but based on those harmed by the defendant's commercial activities. Changing the basis of the distinction recognizes that the equities of the situation may have changed in the presence of unsupervised, autonomous, and thus unpredictable algorithms, and in consideration of who developed (or commissioned), introduced into society the risk of, and is profiting from the activities of what might in some circumstances be considered tiny technological agents of chaos, juxtaposed with who bears the risk of harm from those activities. Several scholars have noted the advisability of ensuring that technology companies are encouraged to internalize the costs of harms resulting from their businesses, and thus encourage more careful conduct that minimizes negative

making others bear the cost. But the basic point is the same: A common enterprise theory permits the law to impose joint liability without having to lay bare and grapple with the details of assigning every aspect of wrongdoing to one party or another; it is enough that in pursuit of a common aim the parties engaged in wrongdoing. That principle could be engrafted onto a new, strict liability regime to address the harms that may be visited on humans by intelligent autonomous machines when it is impossible or impracticable to assign fault to a specific person." Vladeck, *supra* note 310 at 149 (footnotes omitted).

[318] "High-stakes harm" might be thought of as equivalent to the interests protected in Article 22 of the European Union's General Data Protection Regulation (GDPR), which generally prohibits subjecting an individual to solely automated decisions (such as decisions made by an autonomous algorithm) that would have "legal or similarly significant effects" on that person.

external risk, lest such costs be externalized onto more vulnerable parties with less control over the risks and fewer resources to bear the costs.[319]

From this perspective, it is not so much that the law arbitrarily raises the liability standard for a defendant based on consequence, but that the *default* position of the law is that the defendant has engaged in risky conduct attracting the modified negligence standard or strict liability, and lowering the standard to the no-fault insurance pool is an *exception* that recognizes little significant harm actually occurred. This means that if and when injured parties sustain high-stakes, or "legal or similarly significant" systemic harm as a result of unsupervised autonomous platform algorithms, they will have access to meaningful legal redress, fulfilling the several important functions of tort law as expressed earlier. At the same time, where no parties were significantly injured or they sustained injuries in a way that did not engage, for instance, their fundamental human rights, a no-fault insurance pool funded by the AI industry as a whole would provide a "liability release valve" in acknowledgement of the benefits of promoting (responsible, human rights-respecting) innovation.

Putting all of the above together, the following table attempts to balance assessment of initially perceivable risk with the level of actual harm that occurred in order to determine the most appropriate standard of liability, taking into account policy considerations such as victim redress and just allocation of risks and costs (particularly where historically marginalized groups are concerned), corporate accountability, and responsible innovation. The table below is not so much formally proposed as it is offered as an experimental starting point for further discussion:

Liability Standards for Platform Algorithm-Facilitated Emergent Systemic Harms (Unsupervised Autonomous Algorithms)

Level of Actual Harm	Level of Apparent Risk	Liability Standard
Low	Low	No-Fault Insurance Pool
Low	Medium	No-Fault Insurance Pool
Medium/High	Low	Modified Negligence
Medium/High	Medium	Modified Negligence
Low/Medium/High	High	Strict Liability

[319] See e.g. Chandler, *supra* note 310 at 230 (footnotes omitted), ("Citron recommends not a negligence but a strict liability standard in relation to harms caused by breaches of data security. She suggests that security breaches are inevitable even with reasonable security measures. In her view, strict liability would cause database custodians to internalize the full costs of the inevitable data security breaches and so discourage the maintenance of databases except where the benefits of doing so outweigh the entire costs of doing so."); Jackson, *supra* note 19 at 60 ("Similarly, a strict liability regime may indirectly push developers of other technologies who seek to capitalize on the economic benefits of implementing Al into their industries towards a similar cost-benefit scenario that is mutually beneficial for industry and society."); and Bertolini, *supra* note 315 at 240 (writing in the context of foreseeable algorithmic harms, but expressing the same justification regarding internalisation of costs: "Overall, it should be stressed that the ability of the producer of robotic applications to be held liable in cases where the potential harm was foreseeable simply forces this party to internalise the costs of its business choices. Therefore, when designing a robot, if a specific risk in the usage of the machine could be anticipated, the producer would be bound to conceive a safety device to prevent or reduce the risk of actual harm resulting from it, and when that was not currently possible the decision to provide the robot nonetheless with the same capacity should lead to the assessment of the duty to compensate for damage.").

B. Applying Modified Negligence to Platform-Facilitated Emergent Systemic Harms

The modified negligence standard suggested here, for low- and medium- risk endeavours that result in significant harm, is taken from Bathaee,[320] combined with further considerations specific to platform-facilitated emergent systemic harms as developed earlier in this paper. Under Bathaee's framework, two adjustments distinguish the modified negligence test from the traditional test.

First, the focal point of the standard of care analysis moves to an earlier point in the chain of causation, further upstream from the harm. The issue is not whether the algorithm itself met a standard of care in its activities, but "whether the AI's creator or user was negligent in *deploying, testing, or operating the AI.* […] The focus in such a case would be on whether the risk of the potentially unlawful conduct was apparent or should have reasonably been addressed with precautions."[321] While this may at first appear to duplicate the initial assessment of risk that led to using the modified negligence test in the first place, the would-be redundancy is addressed by splitting up "was apparent" (which would occur in the determination of the liability standard) and "should have been addressed" (which would occur in the standard of care analysis). The flow of analysis would occur as follows, using the above table:

a) The court assesses the level of harm that occurred, regardless of how it happened. Medium- or high-level harm would have led to the modified negligence standard.
b) The court then determines whether the initial apparent riskiness of the algorithm was low/medium-risk (leading to modified negligence), or high-risk (leading to strict liability).

The assessment of apparent risk at the liability standards stage is only to establish whether there was any perceivable risk at all, and at what level. It does not require examining or making any determinations about the defendant's conduct and decisions and to what extent they tried to mitigate such risk. That assessment of conduct is what occurs at the standard of care stage in the modified negligence analysis.

Nicholas Price similarly suggests (in a medical context) moving the analytical focus upstream, using

> a standard that would require facilities and health care professionals to exercise "due care in *procedurally evaluating and implementing* black-box algorithms." Under this standard of care, facilities and clinicians would have a duty to evaluate black-box algorithms and to validate the algorithmic results. Under this model, health care professionals are responsible for harm if they did not take adequate measures in properly evaluating the black-box AI technologies used in caring for the patient.[322]

In the context of platform algorithm-facilitated emergent systemic harms, this would mean asking if Facebook exhibited a reasonable standard of care in testing and implementing the algorithms used to produce users' newsfeeds, or evaluating if YouTube was negligent in how it optimized and validated its recommendations algorithm, with regard to the systemic harm under consideration, as opposed to

[320] "When the AI is deployed autonomously in less dangerous or mission-critical settings, a vicarious liability rule may be less appropriate. There may be little risk of harm from the AI's error in these circumstances and holding the user or creator of the AI liable regardless of intent or negligence would chill a large swath of desirable AI applications. Instead, in such cases, the negligent principal rule would be more appropriate. Bathaee, *supra* note 19 at 935

[321] Bathaee, *supra* note 19 at 935 (emphasis added).

[322] Sullivan & Schweikart, *supra* note 19 at 164 (emphasis added). Price elaborates, "Providers and facilities should evaluate black-box algorithms for hallmarks of careful development, including independent validation of algorithmic results and the qualifications of the developers. Facilities are best suited to evaluate algorithms at the point of implementation and should ensure that algorithms – as a whole – are high quality according to measurable characteristics. Providers are able to measure the risk associated with a particular intervention and should accordingly measure the level of validation and confidence against the risks entailed." BBM

determining if Facebook or YouTube could have foreseen or prevented specific instances of, for instance, disinformation, ad-based discrimination, or online violence against protected marginalized communities.

The second difference in the modified negligence test is a departure from foreseeability, in recognition of the black-box nature of the algorithms under analysis. According to Bathaee, the test "should turn not on whether the particular harm caused by the AI was reasonably foreseeable, but whether the harm was a foreseeable consequence of deploying black-box AI autonomously. The question is one of *conceivability, not foreseeability* in such settings."[323] In other words, the test "should focus on the possible effects of deploying AI autonomously without understanding how it functions, rather than on the specific ability of the user or creator of the AI to have predicted the injurious effects of the AI's conduct."[324] This requirement may also contribute to closing the platform foreseeability gap, by not relying on decision-makers at platform companies to agree whether or not something is likely to bring marginalized or vulnerable users to harm (foreseeability), but instead, by demanding that decision-makers more thoughtfully contemplate what could possibly go wrong at all (conceivability).[325] It is important, however, that this notion of conceivability also be applied through an intersectional and relational lens, especially given occasions where platform companies' leaders or software developers have stated that they did not even conceive of the probable, let alone the possible, with respect to potential harm to marginalized or vulnerable individuals arising from their products and services. [326]

While the involvement of unsupervised, autonomous algorithms are central to the nature of the platform-facilitated emergent systemic harms discussed in this paper and subjected to the modified negligence test, emergent systemic harm is defined as resulting from the contributing factors of a platform interacting, only one of which is the algorithm(s). The standard of care analysis (and other elements of the negligence test) must thus consider the platform system as a whole, including the other components of the system, and not just the defendant's conduct with respect to the algorithm alone. Examination of these components would also inform the analysis at all stages of the modified negligence test, alongside the algorithm, while relieving the court of having to depend predominantly on an unintelligible, unexplainable, autonomous technological mechanism to make a substantive determination in the case. Whether or not a platform company's initiative appears low, medium, or high risk at the start, and whether or not the company exercised reasonable care in how it tested, implemented, validated, and operated an unsupervised autonomous algorithm, would depend on the context provided by the other components of the platform system and their known interactions with the algorithm(s). As these components would not share the unintelligibility of unsupervised autonomous algorithms, they could be assessed on a reasonable foreseeability standard (i.e., not conceivability).[327]

[323] Bathaee, *supra* note 19 at 938 (emphasis added).

[324] Bathaee, *supra* note 19 at 935-36

[325] In a way, this hearkens back to the initial, pre-*Mustapha* holding from *Assiniboine* based on *The Wagon Mound (No. 2)*: "The test of foreseeability of damage becomes a question of what is possible rather than what is probable." *Assiniboine*, *supra* note 245 at page 613 (CanLII).

[326] See e.g., Superhuman's founder and CEO, Rahul Vohra, stated the following regarding a location-tracking feature in emails, where a read receipt would tell the sender where the recipient was at the time they opened the email, without the recipient's consent: "We did not imagine the potential for misuse." Rahul Vohra, "3/ I am so very sorry for how our read status feature made folks feel. We did not imagine the potential for misuse. Now we are learning and changing." (3 July 2019 at 18:03), online: *Twitter* <https://twitter.com/rahulvohra/status/1146539947655958528>.

[327] These components, to reiterate from earlier, include: the platform's technological affordances; the platform's business model and associated incentives; the platform's user bases (intended, marketed, and actual user bases) and user behaviour; and how much time had passed and how much the platform user base had grown between its beginning and when the systemic harm fully emerged or became reasonably foreseeable.

The element of time and scale, for example, could go to assessing whether the platform company exercised reasonable care, depending on the suddenness with which the systemic harm emerged. A systemic harm that appears across the entire platform all of a sudden with no warning—such as overnight mass suspensions of user accounts according to an unknown common factor based on an unsupervised autonomous algorithm—may not be considered something the platform had an opportunity to prevent or mitigate, so in the absence of circumstances that ought to have put them on notice, all else being equal and provided the testing, validation, and implementation of the algorithm exhibited reasonable care, the company likely would not be have considered to have fallen below the standard of care. In contrast, where a systemic harm emerges over time and over a growing user base, beginning with incidents that look like discrete, one-off harms to individual users in unique cases, but eventually adding up to demonstrable patterns of discriminatory targeting or systemic oppression across the platform,[328] for instance, then the platform may be found to have violated the relevant standard of care if it did not, at some point in between, investigate the initial "aberrations", examine them for systemic dimensions, or try to mitigate future recurrences or prevent the known harms from becoming systemic.[329]

This broader aspect of the standard of care analysis, rooted in the platform's non-algorithm components, with which the algorithm(s) interact(s), is non-exhaustive, and would also be a suitable home for assessing factors such as whether the platform company cultivates diversity and equity among its leadership, technology, and business teams. This factor would be relevant to the extent such practices would impact internal assessments of risk and reasonably foreseeable harms to marginalized and vulnerable individuals as a result of, for instance, certain platform affordances or the platform's business incentives. If the reasonable person is not racist (and similarly understands other forms of discrimination and being vulnerable to them, as argued above), and the reasonable foreseeability element of standard of care turns on sharing the mindset of a reasonable person, then the platform company has a duty to put itself in a position of being able to assess the potential risks and harms of its own products with a critical intersectional eye, in order to have met the standard of care that would relieve it of liability for emergent systemic harm to the marginalized group(s) represented by the plaintiffs.

C. Strict Liability

Strict liability would be reserved for situations where the risk associated with the impugned platform algorithm(s) was perceivably high from the start—regardless of how nominal or severe the actual harm that occurred was. There are valid arguments against applying strict liability to algorithm-facilitated harms,[330] and they are taken into account by attaching strict liability only to platform algorithms that were perceivably high-risk even at the beginning, based on intended purposes, broader context of the other

[328] To be clear, the initial individual harms themselves would not be considered grounds of liability for platform-facilitated emergent systemic harm, unless they collectively already amounted to a form of systemic harm.

[329] Bathaee offers an example of a similar dynamic in the context of investment algorithms: "The financial institution would be liable to an investor who relied to [their] detriment on an inaccurate appraisal value if the financial institution acted unreasonably in relying on the output of the black-box AI. Perhaps the institution ignored an unreasonable result or apparent bias in the AI's input data, or failed to put in place reasonable safeguards or testing regimes. To *continue* to rely on AI that may be making flawed decisions or that is relying on problematic data may be evidence of willful blindness or may arise to the level of recklessness required for scienter." Bathaee, *supra* note 19 at 933-34 (emphasis added).

[330] See e.g., "Part V rejects strict liability as a potential solution because the unpredictability of AI eliminates the positive effects of strict liability. For instance, if the creator or user of AI cannot predict the effects of the AI ex ante, [they] cannot take precautions for the injury inflicted. Strict liability may only deter smaller firms from developing AI because they would risk outsized liability should the AI cause any injury. This would favor established and well-capitalized participants in the field and erect significant barriers to entry and innovation." Bathaee, *supra* note 19 at 896.

platform system components, and conceivable harms. Despite delineating why strict liability should not generally apply to AI across the board, Bathaee (functionally) makes the following exception:

> When the AI operates autonomously in a mission-critical setting or one that has a high possibility of externalizing the risk of failure on others, such as when it is used in a highly interconnected market or to perform a medical procedure, the AI's user or creator should be more broadly liable for injury the AI inflicts, and a vicarious liability rule is appropriate. In such cases, a lack of transparency should not insulate the user or creator of the AI from liability. Instead, the risks of deploying a black-box AI autonomously in such settings should fall on the AI's user or creator because the AI was used notwithstanding its unpredictable and impenetrable decision-making. In such a case, the imposition of vicarious liability would be functionally equivalent to a strict liability regime. […] [W]hen the AI is both autonomous and unsupervised, the sole question will be whether it was reasonable to have deployed such AI at all. The answer may simply be no, which means that the creator or user of the AI would be liable for the AI's effects, even if [they] could not foresee them and did not intend them.[331]

The above excerpt lands with particular weight in the context of harm to historically marginalized communities, who already live with bearing negative externalities in a variety of contexts and lack the means and resources to absorb the risks and costs of corporate ventures over which they have no control. The principles underlying strict liability for high-risk platform algorithms would also be in accordance with the environmental law doctrines of polluter/beneficiary pays and the precautionary principle. Moreover, applying strict liability in this context is supported by an early argument by Allen Linden that foreseeability outlives its usefulness in some circumstances where it can no longer serve as a rational proxy for the equities of the situation, as opposed to deciding a case directly on the equities themselves.[332]

Conclusion

When it comes to platform regulation and addressing contemporary problems brought about by the rise of multinational online platforms, the field sees new recommendations, solutions, frameworks, and policies that predominantly focus on regulatory responses that would restrict the worst excesses of platform-facilitated systemic harms. Some scholars have acknowledged that a private right of action providing redress for harms arising from online platforms may be required to act as a backstop for regulation. This paper has responded to that opening, by offering some preliminary building blocks of one potential legal approach to holding online platform companies accountable for systemic harm to marginalized groups, where such systemic harm has been facilitated by a platform company.

Principally, this paper has established the concept of platform-facilitated emergent systemic harms to marginalized groups, as a form of harm that could be legally actionable in Canadian law. Emergent systemic harm occurs when constituent parts of a digital platform's ecosystem work together to give rise to an unintended systemic harm that goes beyond the some of the platform's constituent parts. Key contributing factors that are likely to be involved in platform-facilitated emergent systemic harm to marginalized communities include: the platform's algorithms, the platform's design and technological affordances, the platform's business model, time and scale of the platform's reach and user base, and user

[331] Bathaee, *supra* note 19 at 935-36.

[332] "Defendants should be more careful in their activities if they know that they are responsible not only for foreseeable injuries but also for unforeseeable ones. It is also better social cost accounting to make the activity that triggers these results bear the entire cost of the accidents it produces. It is terribly difficult administratively to sort out which injuries are foreseeable and which are not. To avoid this complex job, the courts may have decided to reimburse the plaintiff for all the physical and mental consequences of the injury. Foreseeability theory seems to curtail this kind of analysis. Therefore, I say down with foreseeability!" Allen M Linden, "Down with Foreseeability! Of Thin Skulls and Rescuers" (1969) 47:4 Can Bar Rev 545 at 557.

behaviour and human nature, or the platform's users' individualized and instrumental view of their on-platform activities. Recognizing emergent systemic harm to marginalized groups in Canadian law, as legally actionable, would be supported by Canadian human rights and constitutional law protecting systemically oppressed communities defined by characteristics enumerated or considered analogous under section 15 of the *Charter* or listed under statutory human rights codes. Environmental law principles would also support legal protection against emergent systemic harms, where digital platforms themselves are considered a kind of online ecosystem and part-public, part-privatized information environment.

Close scrutiny of specific examples of what appeared to be platform-facilitated emergent systemic harms to marginalized groups, however, resulted in destabilization of the notion of "emergence" in practice. This destabilization occurs where such harms are shown to be only quasi- or pseudo-emergent systemic harms, due to belatedly revealed involvement of earlier knowledge but continued inaction on the platform's part, or otherwise malfunction, malintent, or bad actors at some point in the system. Such harms may have been "unintended" by the platforms, but they were not always unforeseen, and even in cases where they were ostensibly "unforeseen" by the platforms, the harms were by no means unforesee*able* by the standard of slightly broader and more diverse perspectives. This paper proposed integrating the concepts of intersectional and relational foreseeability into the reasonable foreseeability analysis in Canadian negligence law, based on a more contemporary understanding of the reasonable person, to address the "platform foreseeability gap" when it comes to identifying platform-facilitated risks and harms to marginalized and vulnerable communities.

Further building blocks provided in this paper included: justifying why tort law would be a suitable avenue of recourse for addressing platform-facilitated emergent systemic harms to marginalized groups, specifically through negligence liability; analyzing how traditional foreseeability considerations might apply to platform algorithm-facilitated emergent systemic harms; and examining how the foreseeability analysis would operate in context of autonomous unsupervised algorithms leading to truly emergent systemic harms. The latter involves applying a sliding scale standard of tort liability that moves from a no-fault insurance pool, to a modified negligence test, to strict liability.

Future work in this area could include examining the other elements of the negligence test in the context of platform-facilitated emergent systemic harm to a marginalized group (such as duty of care, standard of care, and remoteness in their own respective rights, as well as causation and defences other than remoteness). Furthermore, platform-facilitated emergent systemic harm *to marginalized groups* is only one category of platform-facilitated emergent systemic harms. There are theoretically many such harms that do not apply specifically to marginalized groups but may be considered "emergent systemic harms of general application". Examples include emergent systemic harm to market competition (or would-be competitors), emergent systemic harm to users' digital self-determination, or emergent systemic harm to voting rights or specific democratic institutions, for instance. The sky is the (foreseeable) limit.