Unjust Enrichment by Algorithm

*Ayelet Gordon-Tapiero & Yotam Kaplan**

Abstract

Social media platforms have become enormously powerful, accumulating wealth at an alarming rate and influencing public opinion with unprecedented efficiency. Platforms use algorithms that promote discriminatory, divisive, extreme, and false content. In recent years, content promoted by social media platforms fueled a series of calamities: the spread of disinformation during the COVID-19 pandemic, the January 6th insurrection, and the establishment of dangerous trends among adolescents and children. The platform crisis is here and is showing no signs of abating.

Platform algorithms recommend divisive, hateful, and inflammatory content because such content encourages users to spend more time on the platform, allows platforms to collect more user data, and presents users with more advertisements, generating more revenue. Thus, the most socially harmful algorithms are the most profitable for platforms. This profitability is fueling the current crisis: as long as harmful algorithms remain the most profitable, new catastrophes are sure to come.

This Article argues that any effective legal response to the platform crisis must address the immense profitability of harmful algorithms. These Authors further suggest that this type of legal response is possible through the doctrine of unjust enrichment. This proposal explains the conditions under which platform profits should be considered unjust, and how the doctrine of unjust enrichment allows courts to strip platforms of such ill-gotten gains. This Article breaks new ground in being the first to study the doctrine of unjust enrichment as a remedy to the platform crisis. Rather than prohibit a particular type of content or a specific optimization metric, this proposal targets platforms' financial incentives, forcing them to consider the broad societal impact of their choices. This is a promising legal venue, offering tools that are unavailable through other frameworks. This Article further details the advantages of this proposal, explains its origins in existing doctrine of the law of unjust enrichment, and provides a rich account of its implementation in practice.

---

* Ayelet Gordon-Tapiero is a Postdoctoral Research Fellow at Georgetown University, Initiative on Tech & Society. Yotam Kaplan is an Associate Professor at Bar-Ilan University Law School ("BIU"). For helpful discussions and insightful comments on earlier drafts, we wish to thank Talia Gillis, Katherine Glenn Bass, Katrina Ligett, Kobbi Nissim, Paul Ohm, Gideon Parchomovsky, Ariel Porat, and participants in the BIU Law School Faculty Workshop. We are grateful to the European Research Council for generous financial support under grant 101077050. We also thank Shay Hay and David Jacobs for excellent research assistance.

TABLE OF CONTENTS

> [W]e don't want to accept/profit from human exploitation.
> –Internal Facebook memo[1]

Introduction

Content personalization on social media is generating immense societal harms.[2] Recently, the spread of disinformation regarding COVID-19 vaccines caused substantial and dangerous vaccine hesitancy.[3] Claims that the dangers of the pandemic were being overstated,[4] along with bogus cures and arguments that the government and the media were exaggerating the severity of the situation, spread on social media like wildfire.[5] Even U.S. President Joe Biden acknowledged that the disinformation spread on social media platforms was "killing people."[6] Despite this, with a global pandemic raging and claiming the

---

1 *See* Justin Scheck, Newley Purnell & Jeff Horwitz, *Facebook Employees Flag Drug Cartels and Human Traffickers. The Company's Response Is Weak, Documents Show.*, Wall St. J. (Sept. 16, 2021, 1:24 PM) (quoting an internal Facebook memo), https://www.wsj.com/articles/facebook-drug-cartels-human-traffickers-response-is-weak-documents-11631812953 [https://perma.cc/H7RR-9KEM].

2 *See* Ayelet Gordon-Tapiero, Alexandra Wood & Katrina Ligett, *The Case for Establishing a Collective Perspective to Address the Harms of Platform Personalization*, 25 Vand. J. Ent. & Tech. L. 635, 651–52 (2023).

3 Neha Puri, Eric A. Coomes, Hourmazd Haghbayan & Keith Gunaratne, *Social Media and Vaccine Hesitancy: New Updates for the Era of COVID-19 and Globalized Infectious Diseases*, 16 Hum. Vaccines & Immunotherapeutics 2586, 2586 (2020) ("As access to technology has improved, social media has attained global penetrance. In contrast to traditional media, social media allow individuals to rapidly create and share content globally without editorial oversight. Users may self-select content streams, contributing to ideological isolation. As such, there are considerable public health concerns raised by antivaccination messaging on such platforms and the consequent potential for downstream vaccine hesitancy, including the compromise of public confidence in future vaccine development . . . .").

4 *See* Ariadne Neureiter, Marlis Stubenvoll, Ruta Kaskeleviciute & Jörg Matthes, *Trust in Science, Perceived Media Exaggeration About COVID-19, and Social Distancing Behavior*, Frontiers Pub. Health, Dec. 1, 2021, at 1, 1 (describing public sentiment that the media was exaggerating the effects and dangers of COVID-19); *see also* Jemma Crew, *Study Reveals One Third of UK Adults Believe Government Is 'Exaggerating' COVID Deaths*, Scotsman (June 1, 2022, 4:55 AM), https://www.scotsman.com/health/study-reveals-one-third-of-uk-adults-believe-government-is-exaggerating-covid-deaths-3715933 [https://perma.cc/42G5-WPKD]; Sofia Bratu, *Threat Perceptions of COVID-19 Pandemic: News Discernment, Media Exaggeration, and Misleading Information*, 19 Analysis & Metaphysics 38, 42 (2020).

5 *See* Alaa Ghoneim, Saiful Salihudin, Isra Thange, Anne Wen, Jan Oledan & Jacob N. Shapiro, *Profiting from Panic: The Bizarre Bogus Cures and Scams of the Coronavirus Era*, Bull. Atomic Scientists (July 24, 2020), https://thebulletin.org/2020/07/profiting-from-panic-the-bizarre-bogus-cures-and-scams-of-the-coronavirus-era/ [https://perma.cc/L4LL-RA4K].

6 Zolan Kanno-Youngs & Cecilia King, *'They're Killing People': Biden Denounces Social Media for Vaccine Disinformation*, N.Y. Times (July 19, 2021), https://www.nytimes.com/2021/07/16/us/politics/biden-facebook-social-media-covid.html [https://perma.cc/NH4S-RWP7].

lives of millions, Facebook's personalization algorithm continued recommending anti-vax content to its users.[7]

Platforms' ability to personalize content for their users is exacerbating distrust in democracy and pushing users to adopt increasingly extreme positions.[8] Social media users are presented with content that reinforces their worldviews and continuously pushes them toward extremism.[9] Some platform users may never encounter a person with opposing views, or conduct a meaningful discussion with them over the platform.[10] Recent changes to platforms' optimization metrics do not promote content that would encourage users to question their beliefs or strengthen their arguments.[11] Instead, platforms' algorithms promote hateful, divisive content, incentivizing content creators to create "outrage bait."[12]

One of the central elements of a functioning democracy is the ability to secure the public's trust in the election process. Mistrust in democratic institutions played a large part in generating the sentiment

---

7 *See A Shot in the Dark: Researchers Peer Under the Lid of Facebook's "Black Box," Uncovering How Its Algorithm Accelerates Anti-Vaccine Content*, AVAAZ (July 21, 2021), https://secure.avaaz.org/campaign/en/fb_algorithm_antivaxx/ [https://perma.cc/54XF-MZYU] (finding that Facebook recommended pages promoting antivaccine content to users). On the term "anti-vax," see Staci L. Benoit & Rachel F. Mauldin, *The "Anti-Vax" Movement: A Quantitative Report on Vaccine Beliefs and Knowledge Across Social Media*, 21 BMC PUB. HEALTH, no. 2106, 2021, at 1, 2 ("A vaccine denier or anti-vaxxer will be defined in this study as someone who believes vaccines do not work, are not safe or refuse vaccines for themselves and their children if applicable.").

8 *See* Luke Munn, *Angry by Design: Toxic Communication and Technical Architectures*, 7 HUMANS. & SOC. SCIS. COMMC'NS, no. 53, 2020, at 1, 6 ("Recommending content based on engagement, then, often means promoting incendiary, controversial, or polarizing content."); Joseph B. Bak-Coleman et al., *Stewardship of Global Collective Behavior*, 118 PROC. NAT'L ACAD. SCIS., no. 27, 2021, at 1, 5 (describing how algorithmic decision-making can facilitate and increase polarization, extremism, and inequality).

9 *See Hearing on "Holding Big Tech Accountable: Targeted Reforms to Tech's Legal Immunity" Before Subcomm. on Consumer Prot., Prod. Safety, & Data Sec. of the S. Comm. on Com., Sci. & Transp.*, 117th Cong. 2 (2021) [hereinafter *Hearing*] (statement of Frances Haugen, former Facebook employee) ("The result has been a system that amplifies division, extremism, and polarization—and undermining societies around the world.").

10 *See* Dominic Spohr, *Fake News and Ideological Polarization: Filter Bubbles and Selective Exposure on Social Media*, 34 BUS. INFO. REV. 150, 151–53 (2017) ("The key issue here is that these groups, convinced of the echo that surrounds them with their own views and preconceptions, in a sense loose [sic] the inclination to proactively discuss ideas with people or groups of a different opinion."); Julie E. Cohen, *Tailoring Election Regulation: The Platform Is the Frame*, 4 GEO. L. TECH. REV. 641, 647 (2020) (claiming that social media users are sorted into "opposing tribes").

11 *See* discussion *infra* Section I.B.1 on the development of optimization metrics.

12 *See* The Journal, *The Facebook Files, Part 4: The Outrage Algorithm*, WALL ST. J., at 17:08 (Sept. 18, 2021), https://www.wsj.com/podcasts/the-journal/the-facebook-files-part-4-the-outrage-algorithm/e619fbb7-43b0-485b-877f-18a98ffa773f [https://perma.cc/2T4N-3WEQ] [hereinafter *Facebook Files*].

that led up to the violent storming of the Capitol on January 6, 2021.[13] The roots of other violent events can be found in content recommended to users by social media platforms.[14] Many are now rightfully concerned with this current state of affairs and fearful of what comes next.[15]

Why is it that platforms recommend such harmful content to their users? After all, they are not in the business of undermining democratic governments. No, the reason is far more prosaic. Platforms recommend divisive, hateful, and extreme content because it is profitable for them to do so.[16] Such content encourages users to spend more time interacting with platforms, allowing platforms to collect more user data, and present users with more advertisements, generating more revenue for them.[17]

This Article offers the first systematic attempt to combat the ongoing platform crisis through the law of unjust enrichment. The law of unjust enrichment allows courts to strip wrongdoers of any ill-gotten gains.[18] This legal tool is meant to ensure that misconduct does not pay

---

13 *The January 6 Effect: An Evolution of Hate and Extremism*, Anti-Defamation League, https://www.adl.org/january-6-effect-evolution-hate-and-extremism [https://perma.cc/6G26-E43E] (explaining that conspiracy theories, including those about election fraud and "stolen" elections, motivated the January 6 insurrection).

14 *See, e.g.*, German Lopez, *Pizzagate, the Fake News Conspiracy Theory that Led a Gunman to DC's Comet Ping Pong, Explained*, Vox (Dec. 8, 2016, 11:15 AM), https://www.vox.com/policy-and-politics/2016/12/5/13842258/pizzagate-comet-ping-pong-fake-news [https://perma.cc/99XR-3BYW]; *see also Hearing*, *supra* note 9, at 2 ("In some cases, this dangerous online talk has led to actual violence that harms and even kills people."); Paul Mozur, *A Genocide Incited on Facebook, with Posts from Myanmar's Military*, N.Y. Times (Oct. 15, 2018), https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html [https://perma.cc/H6DG-5BLG].

15 *See* Jonathan Haidt, *Why the Past 10 Years of American Life Have Been Uniquely Stupid*, The Atlantic (Apr. 11, 2022), https://www.theatlantic.com/magazine/archive/2022/05/social-media-democracy-trust-babel/629369/ [https://perma.cc/K7D9-QEEV]; Jonathan Haidt, *Yes, Social Media Really Is Undermining Democracy*, The Atlantic (July 28, 2022), https://www.theatlantic.com/ideas/archive/2022/07/social-media-harm-facebook-meta-response/670975/ [https://perma.cc/XCA5-BC4V]; Scott Simon, Opinion, *After Jan. 6, What's Next for Our Democracy?*, NPR (June 11, 2022, 08:05 AM), https://www.npr.org/2022/06/11/1104333161/opinion-after-jan-6-whats-next-for-our-democracy [https://perma.cc/6LC2-PBH7].

16 *See Hearing*, *supra* note 9, at 2 ("I saw that Facebook repeatedly encountered conflicts between its own profits and our safety. *Facebook consistently resolved those conflicts in favor of its own profits*. The result has been a system that amplifies division, extremism, and polarization—and undermining societies around the world. In some cases, this dangerous online talk has led to actual violence that harms and even kills people. In other cases, their profit optimizing machine is generating self-harm and self-hate—especially for vulnerable groups, like teenage girls." (emphasis added)).

17 A huge percentage of the revenue of leading social media platforms is generated from ads. *See* Salomé Viljoen, *A Relational Theory of Data Governance*, 131 Yale L.J. 573, 588–89 (2021) ("In 2019, Google reported $134.81 billion in advertising revenue out of $160.74 billion in total revenue. In the first quarter of 2020, Facebook's total advertising revenue amounted to $17.44 billion, compared to $297 million in revenue from other streams." (footnote omitted)).

18 Restatement (Third) of Restitution and Unjust Enrichment § 51(4) (Am. L. Inst. 2011) (describing the disgorgement remedy as designed to strip wrongdoers of gains).

and to remove the incentive to act in ways that are harmful to others.[19] This Article argues this is precisely the remedy required in the present context, to remove platforms' incentive to promote harmful content.[20]

These Authors propose applying the doctrine of unjust enrichment to platform personalization in three categories of cases. The first includes cases where the personalization of content amounts to discriminatory treatment. This is the case, for example, when job ads are presented exclusively to members of one gender or when a particular ethnic group is excluded from the presentation of housing ads.[21] This type of discrimination is already illegal and is therefore a good starting point for the application of the doctrine. The second category of harmful personalization this Article identifies is the promotion of extreme, divisive, and false content that contributes to democratic erosion or political violence. Third and finally, this proposal identifies cases where platforms knowingly abuse sensitive groups by presenting them with content to which they display a particular vulnerability.

These types of personalized recommendations generate immense profits for platforms, as they allow platforms to collect more data about users and present them with more ads.[22] As long as such harmful personalization allows platforms to become enriched, there is no reason for them to refrain from it.[23] This proposal identifies the enrichment generated by harmful personalization as unjust. The application of the doctrine of unjust enrichment to the case of harmful platform personalization is in line with the reasoning and rationale of the doctrine, and a natural development of it.[24] This court-enforced doctrine is the proper legal tool to combat harmful personalization. Compared to regulatory agencies or other regulatory bodies, courts can be less susceptible to regulatory capture[25] and are more accessible

---

19 *See* Ofer Grosskopf, *Protection of Competition Rules Via the Law of Restitution*, 79 TEX. L. REV. 1981, 1997–98 (2001) (explaining that stripping wrongdoers of their gains is necessary to remove incentives for wrongdoing).

20 *See* discussion of the proposal *infra* Section II.B.

21 In the United States, the Fair Housing Act, 42 U.S.C. § 3604, prohibits discrimination in advertising for housing opportunities; the Civil Rights Act of 1964, §§ 703–716, 42 U.S.C. §§ 2000e to 2000e-15, prohibits discrimination in job advertisements based on protected characteristics; the Age Discrimination in Employment Act of 1967, §§ 2–12, 14–15, 17, 29 U.S.C. §§ 621–634, prohibits discrimination in advertising of job opportunities on the basis of age.

22 *See* Gordon-Tapiero et al., *supra* note 2, at 647.

23 *See* Roger McNamee, *Facebook Will Not Fix Itself*, TIME (Oct. 7, 2021, 11:35 AM), https://time.com/6104863/facebook-regulation-roger-mcnamee/ [https://perma.cc/XS9F-5NWN].

24 *See infra* Section II.B.

25 Richard A. Posner, *Regulation (Agencies) Versus Litigation (Courts) An Analytical Framework*, *in* REGULATION VS. LITIGATION: PERSPECTIVES FROM ECONOMICS AND LAW 11, 19 (Daniel P. Kessler ed., 2010) ("Agencies are subject to far more intense interest-group pressures than courts. The agency heads are political appointees and their work is closely monitored by congressional committees. The fact that agency members are specialized, and that they are less insulated from

to unorganized citizens.[26] Regulatory agencies operate by adopting a rule and mandating its implementation. Courts, on the other hand, can apply the doctrine on a case-by-case basis, developing the doctrine and the conditions for its application over time. This measure of flexibility is crucial in the ever-changing world of social media platforms.

The problems caused by harmful platform personalization are frightening. Almost fifty percent of U.S. adults report that they get a large part of their news through social media platforms.[27] Thus, platforms have much control over the type of information they present to individuals, and perhaps, even more importantly, the information they will never expose people to. They have the potential to undermine the way people interact with each other, indeed the very basis upon which democratic societies function. This Article identifies a real opportunity to address these harms. By changing platforms' financial incentives, there is viable potential for change. We cannot allow ourselves as a global society to continue expressing concern over the harms of problematic platform personalization while not taking enough action to prevent them.

This Article makes four novel contributions. The first contribution is *conceptual*. The doctrine of unjust enrichment does not focus exclusively on the harms that personalization generates for individuals and for society. In fact, it is often almost impossible to identify a particular individual harmed by personalization, much less to quantify the damage. Instead, the doctrine focuses on the enrichment experienced by the platform in question. This enrichment is much easier to identify and quantify. Focusing on unjust gains enables the creation of an actionable claim. The second contribution is *doctrinal*. This Article discusses the institutional elements necessary to allow the practical implementation of the doctrine of unjust enrichment to a particular

---

the political process than judges are, makes them targets for influence by special-interest groups; hence the term 'regulatory capture.' Historically, the missions of regulatory agencies have often been anticompetitive, as capture theory implies: interest groups seek to influence agencies to insulate the groups' members from competition, as by blocking new entry. Execution of valid regulatory policies is often thwarted by the dependence of regulators on information supplied by the regulated entities and by the perverse incentives created by 'revolving door' behavior."). See, for example, Rajshree Agarwal & Washington Bytes, *Why Amazon Runs Toward Government with HQ2*, Forbes (Nov. 15, 2018, 7:42 AM), https://www.forbes.com/sites/washingtonbytes/2018/11/15/why-amazon-runs-toward-government-with-hq2/?sh=3311f31067a9 [https://perma.cc/RR2X-H7ZJ], for a discussion of the lobbying efforts of companies such as Amazon, Twitter, Facebook, and Apple to impact regulators' policy making.

26 *See* J. Maria Glover, *The Structural Role of Private Enforcement Mechanisms in Public Law*, 53 Wm. & Mary L. Rev. 1137, 1154 (2012).

27 Mason Walker & Katerina Eva Matsa, *News Consumption Across Social Media in 2021*, Pew Rsch. Ctr. (Sept. 20, 2021), https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/ [https://perma.cc/ZP6T-CT9S] ("A little under half (48%) of U.S. adults say they get news from social media 'often' or 'sometimes,' . . . .").

case of unjust enrichment through personalization. The Article also identifies who may be a potential plaintiff in an unjust enrichment claim made against a social media platform. The third contribution is *analytical*. This Article describes three categories of problematic personalization which generate harms not only for particular individuals, but also for society at large. This proposal identifies not only the harm generated by each category, but, more importantly, the unjust behavior carried out by the platform. For each category the Article also identifies the enrichment mechanism. Finally, this Article makes a *normative* contribution. Based on an analysis of the doctrine of unjust enrichment and its application by courts, the Authors argue that it is legally justified to analyze harmful platform personalization through the lens of unjust enrichment. Despite growing recognition of the severity of the harms driven by platform personalization, regulators and researchers have not yet been able to offer a solution that can effectively prevent platforms from becoming enriched at the expense of the public. Moreover, any attempt to regulate away a particular type of harmful behavior—for example prohibiting the use of downstream Meaningful Social Interaction ("MSI")[28] as an optimization metric—could result in platforms making a slight change so that the new regulation does not directly apply to them. In setting a standard by which platforms' behavior must be examined, this proposal focuses on the way that platforms' incentives are shaped.

The Article proceeds as follows. Part I describes the platform crisis. It explains how information collected from users is used to personalize the content presented to them by describing the development of optimization metrics that guide the activity of platforms' personalization algorithms. The Facebook Files, exposed by Frances Haugen,[29] gives exceptional insight into the behind-the-scenes development of Facebook's personalization algorithm's optimization metric: downstream MSI. The documents not only provide factual information about Facebook's activities, but also expose the concerns raised by Facebook workers that show that they were aware of the harms the platform was causing and were deeply concerned about them.[30] In particular, the quote at the beginning of this Article[31] shows that Facebook workers understood the platform was financially benefitting from exploiting its users. This Part also describes the main harms caused

---

28 *See* discussion *infra* Section I.B.

29 *See* Jeff Horwitz, *The Facebook Files*, WALL ST. J. (Oct. 1, 2021), https://www.wsj.com/articles/the-facebook-files-11631713039 [https://perma.cc/LJ7D-7DPF].

30 *See id.*

31 *Supra* note 1 and accompanying text.

by platforms' personalization: discrimination, the spread of disinformation, increased polarization, and ongoing extremism, culminating in an erosion of trust in democracy and its institutions. Part II proposes the Authors' solution—the application of the doctrine of unjust enrichment to harmful platform personalization. This Part reviews the doctrine of unjust enrichment and demonstrates how each of its elements is suited for addressing the gains generated by platforms' damaging personalization processes. This Part details the three categories of harmful personalization that the Article applies the doctrine to: discrimination, the abuse of vulnerable users, and socially harmful personalization undermining trust in democracy. Part III presents the advantages of applying the doctrine of unjust enrichment to harmful platform personalization. It highlights the fact that the doctrine focuses on gains, not on harms, and therefore does not require identifying an injured party. This Part offers tools and guidelines for calculating the level of enrichment and points out the benefit of applying flexible standards and non-bright-line rules to the innovative practice of platform personalization. The Part offers predictions for several steps that platforms may take in response to the adoption of this proposal and how the application of the doctrine may develop in turn to combat these adaptations. The Conclusion expresses the Authors' sincere hope that this proposal will be a constructive tool for stopping the downward spiral society currently faces.

I. The Platform Crisis

Over the past decade, social media platforms such as Facebook, Instagram, X (formerly known as Twitter), TikTok, and YouTube have emerged as a dominant force in our political, economic, and social lives. Social media platforms drive public opinion, replace traditional market environments, and change the way people interact with each other and experience public life. These deep technological and societal changes are shaped by the commercial interests of platforms as profit maximizing firms and by the ability of platforms to use new technologies to optimize their operations and increase their influence and revenues. These processes have led to unprecedented harms in recent years in the form of discrimination, the abuse of vulnerable users by presentation of harmful content, the spread of disinformation, and the erosion of trust in democracy and its institutions. This Part connects these societal ills with the technology driving platform personalization algorithms and tracks the way in which social media platforms' ability to collect and analyze user data is both central to their business models and deeply harmful for both individuals and society at large. This review of the platform crisis sets the stage for this Article's law reform proposal presented in Part II.

A. The Principles of Personalization

Personalization is fundamental to the operation of social media platforms.[32] This Section highlights the bidirectional nature of platform personalization. First, personalization requires data collection along the *outgoing vector* when data flows *from users to the platform*.[33] While social media platforms typically offer their services "free of charge," users effectively pay for platform services by unwittingly allowing platforms to access and control their data.[34] Second, personalization entails the tailoring of content along the *incoming vector*, along which personalized content is presented *by platforms to users*.[35] Coming together, these basic components of personalization allow platforms to utilize user data to personalize the content each user is presented with, to offer highly targeted advertising services, and to maximize the time users spend actively interacting with the platforms. The basic elements of personalization generate immense power in the hands of social media platforms, leading to the creation of "surveillance capitalism" and driving platform profits.[36]

1. Collecting Data and Building a User Profile

Platforms collect user data along the *outgoing vector*, or when data flows from the user to the platform.[37] Such information includes users'

---

32 Some platforms offer a hybrid option: while basic access is free, these platforms offer subscription models to access a premium version of their services. YouTube allows all users to watch videos and receive personalized recommendations for videos on its platform. Users who pay a monthly subscription receive access to commercial free videos. *See YouTube Premium*, YouTube, https://www.youtube.com/premium [https://perma.cc/RB36-FDA3]. Users who subscribe to Spotify also receive ad-free access to content as well as other premium services. *See Spotify Premium*, Spotify, https://www.spotify.com/us/premium/ [https://perma.cc/G5BZ-2ELC]. While the Authors view paid services as part of social media platforms as well, the paid premium versions operate under a somewhat different business model and therefore fall outside the scope of the analysis in this Article.

33 *See* Gordon-Tapiero et al., *supra* note 2, at 644.

34 *See* Priscilla M. Regan, *A Design for Public Trustee and Privacy Protection Regulation*, 44 Seton Hall Legis. J. 487, 495–96 (2020) ("In exchange for 'free' services . . . individuals provide their personal information . . . ."); *see also* Eli Pariser, The Filter Bubble 16 (2011) ("In exchange for the service of filtering, you hand large companies an enormous amount of data about your daily life—much of which you might not trust friends with.").

The business model used by leading social media platforms is different from those used by other types of platforms. Thus, marketplace platforms like Amazon's marketplace, eBay, Uber, and Airbnb usually charge a percentage of the sum of the transaction conducted on them. *See* Lina M. Kahn, *The Separation of Platforms and Commerce*, 119 Colum. L. Rev. 973, 987 (2019) (describing the business model of marketplace platforms).

35 *See* Gordon-Tapiero et al., *supra* note 2, at 646.

36 *See* Shoshana Zuboff, The Age of Surveillance Capitalism 197 (2020).

37 *See* Gordon-Tapiero et al., *supra* note 2, at 644; *see also* Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. Davis L. Rev. 1183, 1185 (2016) (acknowledging the widespread collection of personal data).

online activity within the platform—posting a tweet, responding to a friend's video, sharing a post viewed in a group, or clicking "like" on certain content—as well as digital activity outside the platform.[38] Some platforms also collect data about their users' offline activity, such as their location or voter registration data.[39] Salome Viljoen highlights the relational nature of data as a meaningful source of information for platforms.[40] Viljoen points out that due to the fact that user data is deeply interconnected, platforms can infer even more data about their users than they were explicitly provided with.[41] When one user uploads a picture of a party they went to, the platform is able to learn that other users appearing in the picture attended the same party whether or not these other users were interested in having this type of information shared and whether or not they were even aware of its existence.[42] A kind neighbor may be unaware they have been captured in their neighbor's smart doorbell or that their conversation was monitored by the neighbor's virtual assistant.[43] Platforms can infer highly personal attributes about users, such as gender, political affiliation, level of income and even medical information despite users actively withholding such

---

38 Facebook tracks its users when they sign into third-party services with their Facebook account. It also gathers information about when its users visit a site embedded with the "like" button, even if the user did not click on it. *See* Jonathan R. Mayer & John C. Mitchell, *Third-Party Web Tracking: Policy and Technology*, 2012 IEEE Symp. on Sec. & Priv. 413, 419; Dina Srinivasan, *The Antitrust Case Against Facebook: A Monopolist's Journey Towards Pervasive Surveillance in Spite of Consumers' Preference for Privacy*, 16 Berkeley Bus. L.J. 39, 41 (2019). Google collects information about news articles its users read. *See* Brian X. Chen, *I Downloaded the Information That Facebook Has on Me. Yikes.*, N.Y. Times (Apr. 11, 2018), https://www.nytimes.com/2018/04/11/technology/personaltech/i-downloaded-the-information-that-facebook-has-on-me-yikes.html [https://perma.cc/RB3F-FXF5] ("Google kept a history of many news articles I had read . . . I didn't click on ads for either of these stories, but the search giant logged them because the sites had loaded ads served by Google.").

39 *See, e.g.*, Pauline T. Kim & Sharion Scott, *Discrimination in Online Employment Recruiting*, 63 St. Louis U. L.J. 93, 97 (2018) ("Facebook also purchases information from data brokers to learn about users' offline behavior, including income and spending habits."); *see also* Giridhari Venkatadri, Piotr Sapiezynski, Elissa M. Redmiles, Alan Mislove, Oana Goga, Michelle L. Mazurek & Krishna P. Gummadi, *Auditing Offline Data Brokers via Facebook's Advertising Platform*, 2019 Proc. World Wide Web Conf. 1920, 1920 ("Recently, data brokers and online services have begun partnering together, allowing for the data collected about users online to be linked against data collected offline. This enables online services to provide advertisers with targeting features that concern users' offline information.").

40 *See* Viljoen, *supra* note 17, at 603.

41 *Id.* at 611.

42 *See* Gergely Biczók & Pern Hui Chia, *Interdependent Privacy: Let Me Share Your Data*, Fin. Cryptography & Data Sec., Apr. 2013, at 338, 340 (describing one user tagging another in a photo as an example of the interdependent nature of data online); *see also* Solon Barocas & Karen Levy, *Privacy Dependencies*, 95 Wash. L. Rev. 555, 568 (2020) ("Or perhaps the Observer takes a photo of Alice, knowing that it will capture Bob in the background.").

43 *See* Barocas & Levy, *supra* note 42, at 568.

data.[44] Some platforms even build "shadow profiles" for individuals who have not registered on the platform by analyzing data gathered from other platform users.[45] This is made possible due to the collective, interconnected nature of data, which enables platforms to detect patterns across large groups.[46]

Social media platforms use the data they have collected and analyzed to draw a detailed profile of their users, including information about their personal attributes, social connections, and interests. The richer the data the platforms hold about each user, the more in depth a profile the platforms are able to draw.[47] This detailed profile enables platforms to present users with content tailored specifically for them. The ability to offer a user personalized content based on data generated by the user and by others is one of the central pillars of social media platforms' business model.[48]

2. Creating a Personalized Platform Experience

Based on the large amounts of data collected and analyzed by platforms, the user experience is personalized along the *incoming vector*, whereby platforms present content to users.[49] All content that users view on the platform is personally tailored and presented to each user

---

44 *See* Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 494, 506 (describing how platforms can infer data about individuals even if they did not provide it); E. Fosch-Villaronga, A. Poulsen, R.A. Søraa & B.H.M. Custers, *A Little Bird Told Me Your Gender: Gender Inferences in Social Media*, INFO. PROCESSING & MGMT., May 2021, at 1, 1 (demonstrating that platforms can infer an individual's gender even when they have not provided it); Kristen M. Altenburger & Johan Ugander, *Monophily in Social Networks Introduces Similarity Among Friends-of-Friends*, 2 NATURE HUM. BEHAV. 284, 284 (2018) ("[E]ven if an individual does not disclose private attribute information about themselves (such as their gender, age, race or political affiliation), methods for relational learning can leverage attributes disclosed by that individual's similar friends to possibly predict their private attributes." (footnotes omitted)).

45 For example, when signing up for Facebook Messenger, users permit Facebook to download their entire list of contacts. *See* Chen, *supra* note 38 ("One surprising part of my index file was a section called Contact Info. This contained the 764 names and phone numbers of everyone in my iPhone's address book. Upon closer inspection, it turned out that Facebook had stored my entire phone book because I had uploaded it when setting up Facebook's messaging app, Messenger."). If enough of a person's friends are active on Facebook Messenger, the app can draw a fairly accurate analysis of that person's social circle despite having no contractual connection to that individual.

46 *See* Gordon-Tapiero et al., *supra* note 2, at 647–51; *see also* Viljoen, *supra* note 17, at 573.

47 Chris Jay Hoofnagle & Jan Whittington, *Free: Accounting for the Costs of the Internet's Most Popular Price*, 61 UCLA L. REV. 606, 608–09 (2014) ("The more time the consumer spends using the service and revealing information, the more the service can adjust the product to reveal more information about the consumer and tailor its advertising of products to that consumer's personal information.").

48 *See* Regan, *supra* note 34, at 496 ("The data that companies acquire from their users enables them to refine the services they offer and to offer new or related services.").

49 *See* Gordon-Tapiero et al., *supra* note 2, at 646.

based on their unique profile; not only the content itself is personalized, but also the order in which content is ranked, the timing in which it is presented, and its frequency.[50]

Social media platforms' ability to personalize ads, as well as other content, plays an important role in their business model. Outside of platforms, advertisers must use various proxies to target their desired audience. This can be achieved, for example, by advertising in a particular location, newspaper or magazine, or on the basis of the content of the web page on which the advertisement is presented.[51] While these venues are chosen because there is an increased probability to reach individuals who are likely to find the ad interesting and relevant and hopefully respond to the ads, these methods of advertising also end up reaching many individuals who have no interest in the product or service being advertised.[52] The highly personalized advertising services offered by social media platforms help advertisers cut back on wasted advertising budgets.[53] Based on data, platforms target ads at users likely to find them interesting and relevant, allowing advertisers to maximize the value of their advertising budgets.[54] Moreover, platforms also control *when* ads are presented to users and can present the ad at a time or context when the user is most likely to respond to it.[55] Despite research questioning the increased effectiveness of personalized advertising, it remains a coveted advertising outlet.[56]

3. Keeping Users Engaged

Operating together, the collection of data along the outgoing vector *and* personalized content along the incoming vector is aimed at assuring maximal user engagement with the platform. This allows

---

50 *See id.*

51 *See* Veronica Marotta, Vibhanshu Abhishek & Alessandro Acquisti, *Online Tracking and Publishers' Revenues: An Empirical Analysis* 2 (Working Paper, 2019), https://weis2019.econinfosec.org/wp-content/uploads/sites/6/2019/05/WEIS_2019_paper_38.pdf [https://perma.cc/RPW9-J4G2].

52 *See* Zhinan Gan & Sang-Bing Tsai, *Research on the Optimization Method of Visual Effect of Outdoor Interactive Advertising Assisted by New Media Technology and Big Data Analysis*, Mathematical Probs. in Eng'g, Dec. 15, 2021, at 1, 1–2 (stating that traditional advertising media is less effective because it is geared toward larger audiences, with less segmentation).

53 *See id.*

54 *See* Marotta et al., *supra* note 51.

55 *See* Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove & Aaron Rieke, *Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes*, Proc. ACM on Hum.-Comput. Interaction, Nov. 2019, at 1, 5 ("[P]latforms try to avoid showing ads from the same advertiser repeatedly in quick succession to the same user; thus, the platforms will sometimes disregard bids for recent winners of the same user. *Second*, the platforms often wish to show users relevant ads . . . .").

56 *See* Marotta et al., *supra* note 51, at 1 (finding that personalized advertising increases advertiser's revenue by only about four percent).

platforms to collect even more information, improve the accuracy of user profiles, and thus offer even more accurately personalized content, present even more ads, and so on. In this vicious cycle, there are only two types of players: winners and users.

Platforms try to monopolize their users' time and attention.[57] They do so by using a variety of addictive features in their interface design.[58] These are reportedly highly successful in generating addiction, especially among younger users.[59] Platforms strive to present users with content they are likely to interact with by liking, retweeting, sharing, commenting, or tagging, among other actions. Users who simply scroll through their feed provide the platform with limited insight into their interests and preferences. Such users are also more likely to leave the platform.[60]

---

57 *See, e.g.*, Rabbit Hole, *Four: Headquarters*, N.Y. TIMES (May 7, 2020), https://www.nytimes.com/2020/05/07/podcasts/rabbit-hole-youtube-susan-wojcicki-virus.html [https://perma.cc/2XR7-75BP]. In this podcast, *New York Times* reporter Kevin Roose suggests that within YouTube, "there was sort of this obsession with growth. There was a very strong push to expand the watch time on the platform and that any challenges that were brought to management around that, that these things just weren't given a real hearing." *Id.* at 11:29.

58 *See* Christian Montag, Bernd Lachmann, Marc Herrlich & Katharina Zweig, *Addictive Features of Social Media/Messenger Platforms and Freemium Games Against the Background of Psychological and Economic Theories*, INT'L J. ENV'T RSCH. & PUB. HEALTH, July 23, 2019, at 1, 4; Catherine Price, *Trapped—the Secret Ways Social Media is Built to be Addictive (and What You Can Do to Fight Back)*, BBC: SCI. FOCUS (Oct. 29, 2018, 4:00 AM), https://www.sciencefocus.com/future-technology/trapped-the-secret-ways-social-media-is-built-to-be-addictive-and-what-you-can-do-to-fight-back/ [https://perma.cc/724X-7LJA] (highlighting that users' feeds are ongoing, never coming to a natural stop or break, thus encouraging users to continuously watch as more and more content appears). Notifications that pop up periodically encourage users to repeatedly check their profiles for new updates, likes, or notifications. *See* ADAM ALTER, IRRESISTIBLE: THE RISE OF ADDICTIVE TECHNOLOGY AND THE BUSINESS OF KEEPING US HOOKED 109–12 (2017); *see also* Mattha Busby, *Social Media Copies Gambling Methods 'to Create Psychological Cravings*,' THE GUARDIAN (May 8, 2018, 2:00 PM), https://www.theguardian.com/technology/2018/may/08/social-media-copies-gambling-methods-to-create-psychological-cravings [https://perma.cc/6XG3-C8WC]. This mechanism builds on individuals' fear of missing out and on the body's natural release of dopamine when encountering an experience worth repeating. In nature, dopamine is released in response to rewarding activities such as eating. *See* R.A. Wise & P.-P. Rompre, *Brain Dopamine and Reward*, 40 ANN. REV. PSYCH. 191, 219 (1989); *see also* Ian McKay, *Up In Smoke: Why Regulating Social Media like Big Tobacco Won't Work (Yet!)*, 97 NOTRE DAME L. REV. 1669, 1680 (2022); Jamie Waters, *Constant Craving: How Digital Media Turned Us All into Dopamine Addicts*, THE GUARDIAN (Aug. 22, 2021, 4:00 PM), https://www.theguardian.com/global/2021/aug/22/how-digital-media-turned-us-all-into-dopamine-addicts-and-what-we-can-do-to-break-the-cycle [https://perma.cc/2NQC-3EZ7].

59 *See* Nandakishor Valakunde & Srinath Ravikumar, *Prediction of Addiction to Social Media*, IEEE INT'L CONF. ELEC., COMPUT. & COMMC'N TECHS., 2019, at 1, 1 ("Social media addiction is a huge problem among the youth."); *see also* Abdullah J. Sultan, *Fear of Missing Out and Self-Disclosure on Social Media: the Paradox of Tie Strength and Social Media Addiction Among Young Users*, 22 YOUNG CONSUMERS 555, 556 (2021) ("[T]he likelihood of becoming addicted to these applications is very high giving [sic] the social benefits that these applications provide for young users.").

60 *See Facebook Files*, *supra* note 12.

Thus, increasing the time and level of interaction of users allows platforms to achieve two main goals: it increases the time available to present users with ads, both personalized and generic, and generates more meaningful data for platforms to use in learning more about their users' interests and preferences.

B. Algorithms of Personalization

Personalized content is generated by an algorithm. Therefore, to understand why users are presented with certain content and to anticipate what type of content is likely to be promoted by a platform, it is important to understand what the algorithms' optimization metrics are.

This Section describes the algorithmic optimization mechanisms that Facebook utilizes in its personalization process along the incoming vector. Platforms' algorithms are highly protected trade secrets, and platforms are reluctant to publicly disclose information about them. In October of 2021, however, Frances Haugen, a former Facebook employee, disclosed a trove of internal Facebook documents to *The Wall Street Journal* in what is now known as "The Facebook Files."[61] These documents revealed new information about internal Facebook operations. Much of the material in the following Sections is based on documents exposed as part of the Facebook Files. In cases where relevant information is available, the Authors also give examples regarding the activity of other social media platforms. There is no reason to believe that the operation of social media platforms other than Facebook is substantially different.

1. Algorithmic Optimization

When it was first launched in 2006, Facebook users had to actively search for their friends' profiles in order to see content other than their own profile.[62] A year later Facebook introduced the "like" button and also enabled users to mark an "X" on content they did not want to see more of in the future.[63] This enabled Facebook to tailor the content each user was presented with to their particular preferences. In 2009, content began being ranked based on its popularity, gauged by the number of "likes" a post had.[64] Posts with the most "likes" were ranked higher in

---

61 *See* Horwitz *supra* note 29; Jeff Horwitz, *The Facebook Whistleblower, Frances Haugen, Says She Wants to Fix the Company, Not Harm It*, WALL ST. J. (Oct. 3, 2021, 7:36 PM), https://www.wsj.com/articles/facebook-whistleblower-frances-haugen-says-she-wants-to-fix-the-company-not-harm-it-11633304122 [https://perma.cc/85VG-6TRT].

62 *See Facebook News Feed Algorithm History*, WALLAROO (Mar. 9, 2023), https://wallaroomedia.com/facebook-newsfeed-algorithm-history/ [https://perma.cc/N7VQ-4JV9].

63 *See id.*

64 *Id.*

users' news feeds.[65] In 2014, Facebook began tracking the time users spent on links they accessed through the platforms.[66] If a user left the outside link immediately, Facebook learned that the user did not like that type of content and would rank similar content lower in the user's newsfeed in the future.[67] Similarly, if the platform detected that a user was spending a lot of time interacting with a certain type of content, it would rank it higher in the future.[68] In 2015 Facebook allowed users not only to mark content that they did not wish to see, but also to select content they wanted to be ranked high in their newsfeed—known as "See First."[69]

The year 2017 was not a good one for Facebook.[70] While users were still spending the same amount of time on the platform, they were becoming increasingly more passive, spending more time watching video content but generating less active engagement.[71] Facebook executives were concerned that users would notice the zombie-like state they were slipping into and leave the platform.[72] Facebook wanted to find a way to encourage users to be more active during the time they spent on the platform: to post more original content, to respond to friends' posts, and to share content they found interesting and relevant.[73] In 2018 Facebook publicly announced that it would be making a change to its algorithm aimed at helping users have more "meaningful interactions," thus promoting users' well-being.[74]

2. Meaningful Social Interaction

To promote users' active engagement with the platform, Facebook changed the metrics that its algorithms were optimizing for. MSI was

---

65 *Id.*

66 *Id.*

67 *Id.*

68 *Id.*

69 *Id.*

70 *See* Stephen Maher, *Facebook's Algorithm Comes Under Scrutiny*, CTR. FOR INT'L GOVERNANCE INNOVATION (Oct. 8, 2021), https://www.cigionline.org/articles/facebooks-algorithm-comes-under-scrutiny/ [https://perma.cc/DH6Z-C9EE].

71 *See* Keach Hagey & Jeff Horwitz, *Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead*, WALL ST. J. (Sept. 15, 2021, 9:26 AM), https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215 [https://perma.cc/FU3D-773U]; *see also* Rachel Metz, *Likes, Anger Emojis and RSVPs: The Math Behind Facebook's News Feed—and How it Backfired*, CNN BUS. (Oct. 27, 2021, 9:51 AM), https://edition.cnn.com/2021/10/27/tech/facebook-papers-meaningful-social-interaction-news-feed-math/index.html [https://perma.cc/UPN6-VHWL] (depicting a document titled "Pre-MSI Trends: Engagement Was Broadly Declining Until 2018H1" as redacted for Congress, which reports a decline in reshares starting in 2017).

72 Hagey & Horwitz, *supra* note 71.

73 *Id.*

74 Adam Mosseri, *News Feed FYI: Bringing People Closer Together*, META (Jan. 11, 2018), https://www.facebook.com/business/news/news-feed-fyi-bringing-people-closer-together [https://perma.cc/L23F-3KHU].

selected as the new optimization criteria.[75] MSI consists of two parameters: the number of interactions content receives across users and how close the users interacting with the content are with each other. Content created by a close connection—for example, somebody with many friends in common with the user in question—and content that generated a high level of engagement across users received a higher MSI score.[76] Presenting users with more content created by close friends and family and less content created by businesses and media was expected to increase users' engagement with the platform while also increasing their well-being.[77]

A post's MSI ranking is determined by summing up the value of each user's interaction with it. Thus, a like by any single user is worth one point, while a reaction emoji or a reshare generates five points.[78] More significant engagement, such as commenting, adds another thirty points to the post's score.[79] A post's MSI score is generated per *post* and not for each *user*.[80] Facebook determined that the level of engagement was to be given more weight compared to closeness in determining a post's MSI score.[81] Thus, the overall number of points generated by the various engagements with a particular post was multiplied by a number that was supposed to serve as a proxy for the level of closeness between the people interacting.[82] For an interaction with a close friend, the engagement score would be multiplied by 0.5, while interaction with a complete stranger would be multiplied only by 0.3.[83] The result of this calculation was the content's MSI score for a specific user.[84]

Content with a high MSI score would be promoted and shown to users in their newsfeed based on the expectation that they too were likely to find it interesting and to interact with it.[85] At the same time, content with a low MSI score would not be promoted by the platform as it was deemed uninteresting and unlikely to encourage engagement.[86] Indeed, switching optimization metrics from optimizing for likes or emojis to optimizing for engagement generated a newsfeed that users

---

75 Hagey & Horwitz, *supra* note 71.

76 *See Facebook Files*, *supra* note 12, at 10:00.

77 *See* Seth Fiegerman & Laurie Segall, *Facebook to Show More Content from Friends, Less from Publishers and Brands*, CNN Bus. (Jan. 11, 2018, 8:41 PM), https://money.cnn.com/2018/01/11/technology/facebook-news-feed-change/index.html [https://perma.cc/6DU3-Q49F].

78 Metz, *supra* note 71.

79 *Id.*

80 *See id.*

81 *See id.*

82 *See id.*

83 *See id.*

84 *See id.*

85 *See id.*

86 *See id.*

were interacting with more.[87] When announcing these changes to the algorithm's optimization metrics, Mark Zuckerberg, Facebook CEO, stated, "By making these changes, I expect the time people spend on Facebook and some measures of engagement will go down. But I also expect the time you do spend on Facebook will be more valuable."[88]

Reality was much removed from this expectation. Contrary to Facebook's public position, there was absolutely no sign that this change had been successful in increasing users' well-being or making the time users spent on the platform more valuable to them.[89] If anything, the opposite was true. The MSI criterion did not ask what and whose content users would *enjoy* seeing and interacting with. Rather, this optimization metric effectively asked what content would *elicit a response* from users. Unfortunately, it turned out that content with a high MSI score was not high-quality, thought-provoking, dialogue-encouraging content but rather tended to be content that was negative, outrageous, toxic, and divisive.[90] Content creators, who wanted their content to continue being promoted and reach a broad audience, were therefore incentivized to generate "outrage bait" to increase the likelihood of their content being promoted by the algorithm.[91] Facebook prioritized content that sparked controversy, not well-being.

Internal Facebook documents show that the company was well aware of the detrimental effects optimizing for MSI had on the type of content that was being promoted. An internal Facebook memo from 2018 titled, "Does Facebook reward outrage? Posts that generate negative comments get more clicks," reported that angry comments on content posted by *BuzzFeed* led to more engagement with it.[92] In another internal post, a Facebook employee reported that "[p]olitical parties across Europe claim that Facebook's algorithm change in 2018 . . . has changed the nature of politics. For the worse."[93]

---

87 *See Facebook Files*, *supra* note 12, at 10:55.

88 Fiegerman & Segall, *supra* note 77.

89 *See Facebook Files*, *supra* note 12, at 11:20.

90 *See* Keith Zubrow, Maria Gavrilovic & Alex Ortiz, *Whistleblower's SEC Complaint: Facebook Knew Platform was Used to "Promote Human Trafficking and Domestic Servitude*," CBS NEWS (Oct. 4, 2021, 6:16 PM), https://www.cbsnews.com/news/facebook-whistleblower-sec-complaint-60-minutes-2021-10-04/ [https://perma.cc/D7VX-XQ27].

91 *See Facebook Files*, *supra* note 12, at 17:08; *see also* Metz, *supra* note 71.

92 Metz, *supra* note 71.

93 David Ingram, Olivia Solon, Brandy Zadrozny & Cyrus Farivar, *The Facebook Papers: Documents Reveal Internal Fury and Dissent over Site's Policies*, NBC NEWS (Oct. 25, 2021, 2:16 PM) (quoting a Facebook employee), https://www.nbcnews.com/tech/tech-news/facebook-whistleblower-documents-detail-deep-look-facebook-rcna3580 [https://perma.cc/M3GN-RQ59]. Scott Simms, a Canadian parliament member expressed similar sentiments. *See* Maher, *supra* note 70.

3. Downstream MSI

Throughout the first year during which MSI was implemented, Facebook found that it was successful in increasing most types of user engagement.[94] Next in the development of its optimization metrics, Facebook began rating content based on the *expected* engagement it was anticipated to generate and called this new optimization metric "downstream MSI."[95] This new metric was made possible because the platform had analyzed the type of content that ranked high on the MSI score in the past and became confident in its ability to predict the type of content that would generate high MSI scores in the future.[96]

Optimizing for downstream MSI further increased the promotion of toxic, divisive, and polarizing content. Users were indeed more likely to engage with content with a high downstream MSI score, for example, by fighting with each other in the comments section.[97] The Facebook Files revealed that Facebook was aware that downstream MSI was facilitating the promotion of even more harmful, divisive, conspiratorial, and fake content.[98]

The Civic Integrity Team at Facebook, a team whose task was to combat hate speech and disinformation on the platform,[99] expressed concern over the type of content that was being promoted by adopting downstream MSI as the optimization metric. They highlighted that the content being promoted was often harmful, negative, divisive, and false.[100] While Facebook implemented suggestions the team made on how to slow the spread of harmful content, as well as potential changes to the algorithm in certain countries and in particular contexts, they were not broadly implemented across the platform.[101] Facebook was concerned that slowing down the virality of content would lower user engagement and adversely affect the platforms' income.[102]

During the same period of time, YouTube also started making changes to its recommendation algorithm. Rather than just presenting users with random new videos or creators, YouTube wanted to be in a position to recommend content to a user that they would find interesting, even before that user themselves knew that they would likely

---

94 *See* Metz, *supra* note 71.

95 *See* Hagey & Horwitz, *supra* note 71.

96 *See Facebook Files*, *supra* note 12, at 15:50.

97 *Id.*

98 *Id.* at 16:55.

99 *Id.* at 18:55.

100 *Id.* at 11:50.

101 *See id.* at 20:30; *see also* Mozur, *supra* note 14 (explaining the larger political impact of Facebook on Myanmar).

102 *Facebook Files*, *supra* note 12, at 22:50.

be interested in a particular type of video.[103] The change in YouTube's algorithm resulted in the promotion of more extreme and polarizing views.[104] In an interview with *New York Times* reporters Kevin Roose and Andy Mills, former YouTube CEO, Susan Wojcicki, acknowledged that YouTube was concerned about the state of their recommendation algorithm and the content it was promoting.[105]

Ultimately, optimizing for downstream MSI achieves platforms' goal of increasing user engagement. There is no denying the fact that divisive, inflammatory content encourages users to spend more time generating engagement and traffic for platforms.[106] This increased activity allows platforms to collect more data points for each user and provides them with a better ability and opportunity to increasingly fine tune the ads presented to these users. Presenting the right user with the ad they are most likely to respond to at a time when they are most likely to be susceptible to the content of the ad is what drives these platforms' business model.

C. The Harms of Personalization

Personalization has now become ubiquitous. Local supermarkets send personalized coupons,[107] and navigation apps show users restaurants that they might like as users approach them.[108] Netflix offers recommended shows based on past choices,[109] and Spotify can introduce users to new artists they are likely to enjoy.[110] Personalization online as well as offline has become part of everyday life and offers multiple

---

103 In Rabbit Hole, *supra* note 57, at 07:28, Susan Wojcicki, former CEO of YouTube, explained that understanding what people will be interested in is "the hardest area for us to discover. Interests that you haven't necessarily told us that you're interested in . . . or you might not know . . . . [W]e certainly have gotten better at predicting what people are interested in."

104 *See generally id.*

105 *Id.* at 9:30.

106 *See Facebook Files*, *supra* note 12, at 18:55.

107 These can sometimes cause embarrassing results. In 2012, retail giant Target sent one of their young shoppers coupons for baby-related products after their "'pregnancy prediction' score" determined she was pregnant. Her father found out about his daughter's pregnancy after seeing the coupons. *See* Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. Times Mag. (Feb. 16, 2012), https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html [https://perma.cc/RCG8-KSAT].

108 *See, e.g.*, Shelby Brown, *7 Google Features to Use When You Don't Know What's for Dinner*, CNET (Nov. 28, 2022, 4:00 AM), https://www.cnet.com/tech/services-and-software/7-google-feature-to-use-when-you-dont-know-whats-for-dinner/ [https://perma.cc/TS3Q-CBDC].

109 *How Netflix's Recommendations System Works*, Netflix, https://help.netflix.com/en/node/100639 [https://perma.cc/6JNW-PWCY].

110 Charlotte Hu, *Why Spotify's Music Recommendations Always Seem So Spot On*, Popular Sci. (Dec. 2, 2021, 8:00 PM), https://www.popsci.com/technology/spotify-audio-recommendation-research/ [https://perma.cc/B6RL-8B5K]; Nick Seaver, Computing Taste: Algorithms and the Makers of Music Recommendation 49–71 (2022).

benefits.[111] At the same time, personalization is also a form of manipulation. Platforms use personalization to constantly try to nudge their users to act in ways that will benefit the platform. Not all manipulative platform behavior is cause for the same level of concern.[112] While suggesting a user wish their mother happy birthday seems to rank low on a scale of manipulative behavior, experimenting with users' emotions ranks high on this scale and compromises users' autonomy.[113]

Evidence accumulated over recent years demonstrates that, in practice, platforms' ability to manipulate their users through content personalization gives rise to a variety of harms. These harms, extensively researched and analyzed in the literature, are reviewed below.

1. Discrimination

Systematic personalization of content may result in illegal discrimination.[114] Discriminatory personalization is particularly relevant in the context of ads for jobs and housing opportunities where it is already recognized as illegal.[115] Even differential personalization that does not amount to strictly illegal discrimination can be harmful. If teenage boys on social media are presented with content about the recent scientific discoveries of the Webb telescope while girls are presented with makeup tutorials, this could be viewed as harmful and risks perpetuating gender biases, even though it is not currently illegal. It is similarly

---

111 *See* David Doty, *A Reality Check on Advertising Relevancy and Personalization*, Forbes (Aug. 13, 2019, 12:51 PM), https://www.forbes.com/sites/daviddoty/2019/08/13/a-reality-check-on-advertising-relevancy-and-personalization/?sh=24abc3837690 [https://perma.cc/VAL9-FNMV].

112 *See* T. M. Wilkinson, *Nudging and Manipulation*, 61 Pol. Stud. 341, 342 (2013) (recognizing that there are different levels of manipulation); *see also* Yochai Benkler, The Wealth of Networks: How Social Production Transforms Markets and Freedom 141 (2006) ("We experience some decisions as being more free than others . . . .").

113 *See generally* Adam D.I. Kramer, Jamie E. Guillory & Jeffrey T. Hancock, *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks*, 111 Proc. Nat'l Acad. Scis. 8788 (2014) (reporting the experiment and its outcomes). *See also* Evan Selinger & Woodrow Hartzog, *Facebook's Emotional Contagion Study and the Ethical Problem of Co-opted Identity in Mediated Environments Where Users Lack Control*, 12 Rsch. Ethics 35, 35 (2016) (highlighting the problematic aspects of the Facebook experiment); Yochai Benkler, *Degrees of Freedom, Dimensions of Power*, 145 Daedalus 18, 23 (2016) (giving the Facebook experiment as an example of the power platforms wield over their users).

114 While discriminating based on gender in the context of job and housing opportunities is illegal, there are other contexts where differential treatment based on protected attributes is not considered illegal discrimination, though this does not mean that it should be allowed. On the disparate impact of algorithms see generally Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI*, Comput. L. & Sec. Rev. Mar. 2020, at 1, 64. *See also* Sigal Samuel, *Why It's So Damn Hard to Make AI Fair and Unbiased*, Vox (Apr. 19, 2022, 6:00 AM), https://www.vox.com/future-perfect/22916602/ai-bias-fairness-tradeoffs-artificial-intelligence [https://perma.cc/T3DQ-YVVV].

115 *See* sources cited *supra* note 21.

wrong if members of one race are presented with advertisements for beer and fast food while members of another race are presented with content promoting healthy eating. This could contribute to disturbing health disparities.[116] Differential treatment based on protected attributes harms both the particular individual being targeted and society at large.

2. Disinformation

In recent years, the intentional spread of disinformation has expanded and caused increased concern.[117] The term "disinformation" is used to describe content which is fake, purposely misleading, and manipulative.[118] It also includes content generated by an imposter claiming to be a reliable source.[119] Disinformation is yet another form of manipulation as it attempts to overcome individuals' judgment, tricking them to believe false content, confusing them about what is real or what source they can rely on, and making them generally doubtful and unbelieving.[120] While the spread of disinformation is, by its very nature, harmful and deceitful, its effects are substantially exacerbated by platforms' ability to personalize content for different users: platforms are financially incentivized to determine precisely which users are more

---

116 *See* Jennifer L. Harris & Willie Frazier III, Rudd Ctr. for Food Pol'y & Obesity, Increasing Disparities in Unhealthy Food Advertising Targeted to Hispanic and Black Families 1, 6–8, 11 (2019).

117 Edson C. Tandoc Jr., Zheng Wei Lim & Richard Ling, *Defining "Fake News" A Typology of Scholarly Definitions*, 6 Digit. Journalism 137, 139 (2018) (providing a typology of types of fake news).

118 *See id.* at 140.

119 *See* Eleni Kapantai, Androniki Christopoulou, Christos Berberidis & Vassilios Peristeras, *A Systematic Literature Review on Disinformation: Toward a Unified Taxonomical Framework*, New Media & Soc'y, 2020 at 1, 23; *see also* David M.J. Lazer et al., *The Science of Fake News*, 359 Science 1094, 1094 (2018) ("Fake news has primarily drawn recent attention in a political context but it also has been documented in information promulgated about topics such as vaccination, nutrition, and stock values."); Gilad Lotan, *Fake News Is Not the Only Problem*, Points (Nov. 23, 2016), https://medium.com/datasociety-points/fake-news-is-not-the-problem-f00ec8cdfcb [https://perma.cc/JPD4-2KWE] ("Biased information—misleading in nature, typically used to promote or publicize a particular political cause or point of view—is a much more prevalent problem than fake news.").

120 *See* Yochai Benkler, Casey Tilton, Bruce Etling, Hal Roberts, Justin Clark, Robert Faris, Jonas Kaiser & Carolyn Schmitt, Berkman Klein Ctr., Mail-In Voter Fraud: Anatomy of a Disinformation Campaign 2–3 (2020) (discussing how fake news also facilitates distrust in democracy and basic democratic processes such as elections, pointing to the U.S. presidential elections as an example); *see also* Mallory Newall, *More than 1 in 3 Americans Believe a 'Deep State' Is Working to Undermine Trump*, Ipsos (Dec. 30, 2020), https://www.ipsos.com/en-us/news-polls/npr-misinformation-123020 [https://perma.cc/S57R-CZX9] ("[F]ewer than half (47%) are able to correctly identify that this statement is false: 'A group of Satan-worshipping elites who run a child sex ring are trying to control our politics and media.' Thirty-seven percent are unsure whether this theory backed by QAnon is true or false, and 17% believe it to be true.").

likely to engage with specific types of disinformation and can then present them with such content. Targeting disinformation at susceptible users massively amplifies its spread and impact and generates more data and revenue for platforms.[121]

3. Extremism and Polarization

Personalized content on social media platforms becomes increasingly extreme and polarizing over time.[122] This personalization process can encourage a radicalization of thought processes. For example, an individual who shows an initial propensity to conspiracy theories can expect to be presented with a growing number of recommendations for content reaffirming and accentuating such beliefs.[123] Users presented with conspiracy theories online have carried out violent acts offline based on their belief of such theories that platforms continuously

---

121 *See* Peter Cohan, *Does Facebook Generate Over Half of Its Ad Revenue from Fake News?*, FORBES (Nov. 25, 2016, 10:36 AM), https://www.forbes.com/sites/petercohan/2016/11/25/does-facebook-generate-over-half-its-revenue-from-fake-news/?sh=27c547f3375f [https://perma.cc/7JDS-QT76].

122 Zeynep Tufekci, *YouTube, the Great Radicalizer*, N.Y. TIMES (Mar. 10, 2018), https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html [https://perma.cc/MB9J-YR9V] ("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."); Jeff Horwitz & Deepa Seetharaman, *Facebook Executives Shut Down Efforts to Make the Site Less Divisive*, WALL ST. J. (May 26, 2020, 11:38 AM), https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499?mod=hp_lead_pos5 [https://perma.cc/9WM3-ABUT] ("Our algorithms exploit the human brain's attraction to divisiveness. . . . If left unchecked [Facebook's users would be presented with] more and more divisive content in an effort to gain user attention & increase time on the platform." (quoting a 2018 Facebook presentation)).

123 *See* Panagiotis Metaxas & Samantha Finn, *The Infamous #Pizzagate Conspiracy Theory: Insight from a TwitterTrails Investigation*, WELLESLEY COLL. FAC. SCHOLARSHIP, 2017, at 1, 4 (arguing that echo chambers, promoted by platforms, create a perfect environment for the spreading of conspiracy theories); Brandy Zadrozny, *Fire at 'Pizzagate' Shop Reignites Conspiracy Theorists Who Find a Home on Facebook*, NBC NEWS (Feb. 1, 2019, 5:55 PM), https://www.nbcnews.com/tech/social-media/fire-pizzagate-shop-reignites-conspiracy-theorists-who-find-home-facebook-n965956 [https://perma.cc/7PEA-NM3S] ("In the case of conspiracy content, Facebook's recommendation engine says, 'If you like pseudoscience, I'll show you chemtrails and flat earth.'" (quoting Renée DiResta, director of research at New Knowledge)); *see, e.g.*, Brandy Zadrozny & Ben Collins, *How Three Conspiracy Theorists Took 'Q' and Sparked Qanon*, NBC NEWS (Aug. 14, 2018, 12:25 PM), https://www.nbcnews.com/tech/tech-news/how-three-conspiracy-theorists-took-q-sparked-qanon-n900531 [https://perma.cc/RA8Z-ML6G] ("There are now dozens of commentators who dissect "Q" posts . . . but the theory was first championed by a handful of people who worked together to stir discussion of the "Q" posts, eventually pushing the theory on to bigger platforms and gaining followers—a strategy that proved to be the key to Qanon's spread and the originators' financial gain.").

presented to them.[124] Individuals that have different starting points and are pushed in opposite extremes become increasingly polarized.[125] Presenting users with extreme content keeps them engaged and generates ongoing profits for platforms.[126]

4. Democratic Erosion

The various types of harms described above raise special concerns regarding their impact on political discourse and democratic stability worldwide.[127] In the days and hours leading up to the January 6th storming of the Capitol, social media platforms became a stage for QAnon.[128] Facebook's decision to dissolve the civic integrity team after the 2020 elections has been cited as a decision that made it harder for the platform to identify and prevent the spread of harmful content that contributed to the January 6th insurrection.[129]

The fact that people are not exposed to opinions different from their own thwarts their ability to form a perception of the distribution of public opinion that would be "likely to promote a sense of legitimacy

---

124 *Man Pleads Guilty to Setting Fire at 'Pizzagate' Restaurant in D.C.*, NBC NEWS (Dec. 18, 2019, 8:00 AM), https://www.nbcnews.com/news/us-news/man-pleads-guilty-setting-fire-pizzagate-restaurant-d-c-n1103691 [https://perma.cc/T3ZD-7XJA] (describing a violent attack carried out by a conspiracy theory believer).

125 Moran Yarchi, Christian Baden & Neta Kligler-Vilenchik, *Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media*, 38 POL. COMMC'N 98 (2021) (explaining that interactional polarization "focuses on a process whereby participants in a debate increasingly interact with like-minded individuals, while disengaging from interactions with others who hold opposing viewpoints").

126 *See* Lina M. Kahn & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497, 505 (2019) ("Divisive and inflammatory content is good for business."); John Naughton, *Extremism Pays. That's Why Silicon Valley Isn't Shutting It Down*, THE GUARDIAN (Mar. 18, 2018, 3:00 AM), https://www.theguardian.com/commentisfree/2018/mar/18/extremism-pays-why-silicon-valley-not-shutting-it-down-youtube [https://perma.cc/HB4E-BW75] ("[U]nderpinning the implicit logic of [YouTube's] recommender algorithms is evidence that people are drawn to content that is more extreme than what they started with . . . .").

127 Democracy and democratic stability have been on the decline globally for over a decade. *See generally* Larry Diamond, *The Democratic Rollback: The Resurgence of the Predatory State*, 87 FOREIGN AFFS. 36 (2008); Arch Puddington, *The 2008 Freedom House Survey: A Third Year of Decline*, 20 J. DEMOCRACY 93 (2009); Arch Puddington, *The Freedom House Survey for 2009: The Erosion Accelerates*, 21 J. DEMOCRACY 136 (2010); Joshua Kurlantzick, *The Great Democracy Meltdown*, NEW REPUBLIC (May 19, 2011), https://newrepublic.com/article/88632/failing-democracy-venezuela-arab-spring [https://perma.cc/UE2Y-Z3LQ].

128 *See* Craig Timberg, Elizabeth Dwoskin & Reed Albergotti, *Inside Facebook, Jan. 6 Violence Fueled Anger, Regret over Missed Warning Signs*, WASH. POST (Oct. 22, 2021, 7:36 PM), https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/ [https://perma.cc/Y3H6-WEGZ].

129 Billy Perrigo, *How Facebook Forced a Reckoning by Shutting Down the Team that Put People Ahead of Profits*, TIME (Oct. 7, 2021, 11:35 AM), https://time.com/6104899/facebook-reckoning-frances-haugen/ [https://perma.cc/7629-GL6Z].

for democratic outcomes, [an increase in] people's ability to generate reasons for their political opinions and their ability to differentiate among ideologically distinct attitudes, and a stimulus effect on political participation."[130] Thus, polarization can decrease trust in elected officials, in democratic institutions, and, more generally, in democracy as a legitimate form of government.[131] The storming of the Capitol was another frightening reminder that words online can translate into real world violence.

Frances Haugen, the Facebook whistleblower, sums up the central challenge in the context of these harms:

> The thing I saw at Facebook over and over again was there were conflicts of interest between what was good for the public and what was good for Facebook. And Facebook, over and over again, chose to optimize for its own interests like making more money. . . . [Facebook's] incentives are misaligned, right? Like, Facebook makes more money when you consume more content. . . . [O]ne of the consequences of how Facebook is picking out that content today is it is—optimizing for content that gets engagement, or reaction. But its own research is showing that content that is hateful, that is divisive, that is polarizing, it's easier to inspire people to anger than it is to other emotions.[132]

To sum, the current freedom given to platforms to collect data and personalize content in a way that serves their financial purposes is generating immense societal harm.

II. Platform Personalization as Unjust Enrichment

This Part explores the possibility of contending with the platform crisis through the conceptual framework offered by the law of unjust enrichment. Under the current structure of platform personalization, certain elements of platform revenues should be considered unjust enrichment and thus be subject to restitution. It starts by offering a general overview of the relevant elements of the law of unjust

---

130 Dominic Spohr, *Fake News and Ideological Polarization: Filter Bubbles and Selective Exposure on Social Media*, 34 Bus. Info. Rev. 150, 152 (2017) (alteration in original) (quoting J Brundidge, *Encountering "Difference" in the Contemporary Public Sphere*, 60 J. Commc'n 680 (2010)).

131 *See* Julie E. Cohen, *Tailoring Election Regulation: The Platform Is the Frame*, 4 Geo. L. Tech. Rev. 641, 659 (2020) (suggesting that the way that public discourse occurs on platforms undermines the structure necessary for a stable democracy).

132 Scott Pelley, *Whistleblower: Facebook Is Misleading the Public on Progress Against Hate Speech, Violence, Misinformation*, CBS News: 60 Minutes (Oct. 4, 2021, 7:32 AM), https://www.cbsnews.com/news/facebook-whistleblower-frances-haugen-misinformation-public-60-minutes-2021-10-03 [https://perma.cc/9XW9-GWER].

enrichment and then applies them to three categories of harmful platform personalization.

A. The Law of Unjust Enrichment

A person unjustly enriched at the expense of another must make restitution of any undeserved benefits.[133] This is the general maxim of the law of restitution,[134] also known as the law of unjust enrichment.[135] This maxim is considered a basic moral principle, and a fundamental element of the legal system.

The practical legal applications of this general principle are numerous and varied.[136] For instance, a recipient of a mistaken payment is typically considered to have been unjustly enriched at the payer's expense.[137] In such a case, the recipient is made to make restitution of the sums mistakenly received,[138] subject to some defense rules.[139] The law of unjust enrichment also operates when goods or services are provided without a contract.[140] Thus, an individual who received lifesaving treatment in the case of an emergency can be considered to have been unjustly enriched at the expense of the medical services provider.[141] In such a case, the beneficiary is typically obligated to pay fair market price for the services they received, even when no contract was formed between the parties.[142]

---

133 RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT § 1 (AM. LAW INST. 2011) ("A person who is unjustly enriched at the expense of another is subject to liability in restitution."); WARD FARNSWORTH, RESTITUTION: CIVIL LIABILITY FOR UNJUST ENRICHMENT 1–2 (2014).

134 *See* Douglas Laycock, *The Scope and Significance of Restitution*, 67 TEX. L. REV. 1277, 1278 (1989).

135 *Id.*

136 *See* Emily Sherwin, *Restitution and Equity: An Analysis of the Principle of Unjust Enrichment*, 79 TEX. L. REV. 2083, 2108–12 (2001) (noting the multiplicity of legal categories cohabitating under the broad umbrella of "unjust enrichment").

137 PETER BIRKS, UNJUST ENRICHMENT 3 (2d ed. 2005).

138 RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT § 57 (AM. L. INST. 2011); *see also* HANOCH DAGAN, THE LAW AND ETHICS OF RESTITUTION 11–25, 37–85 (2004); Andrew Burrows, *Restitution of Mistaken Enrichments*, 92 B.U. L. REV. 767, 767 (2012); BIRKS, *supra* note 137, at 3.

139 *See* RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT § 65 cmt. a (AM. L. INST. 2011); Andrew Kull, *Defenses to Restitution: The Bona Fide Creditor*, 81 B.U. L. REV. 919, 921–22 (2001). The doctrine of change of position is one central defense in such cases, used to limit restitution when the recipient of a mistaken payment relied on the mistaken payment in good faith, such that returning it to the payer would cause the recipient loss. *See* RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT § 65 cmts. a, d (AM. L. INST. 2011).

140 BIRKS, *supra* note 137, at 39–40.

141 *See id.*

142 *See, e.g.*, K.A.L. v. S. Med. Bus. Servs., 854 So. 2d 106, 107–08 (Ala. Civ. App. 2003) (an unconscious patient was brought to the hospital after a failed suicide attempt; the patient's life was saved and the hospital was entitled to restitution for reasonable costs); *In re* Est. of Boyd, 8 P.3d 664, 669 (Idaho Ct. App. 2000) (a patient was admitted to the hospital by his wife and stepson and refused to pay medical bills; the court granted restitution); *In re* Crisan Est., 107 N.W.2d 907, 910–11

The key normative question in all unjust enrichment cases is the extent to which an enrichment caused an "injustice."[143] Thus, in the mistaken payment case, the recipient's enrichment is considered unjust as it was unintentional.[144] In the emergency medicine case, the beneficiary's enrichment is unjust as it constitutes a windfall and the beneficiary-defendant did not pay for it.[145] In both cases, the defendant can be obligated to pay restitution, and their enrichment is considered "unjust," even though there is no wrongful conduct by the defendant. That is, the defendant did not breach a promise to pay—as they made no promise at all—and similarly did not breach a duty of care and therefore cannot be liable in tort. The defendant committed no crime and violated no regulatory order but can still be liable based on their unjust enrichment.

Yet, in other cases, the defendant's enrichment is considered unjust because it was obtained through the defendant's wrongful conduct or even through the defendant's crime.[146] For instance, in the classic case of *Riggs v. Palmer*,[147] the defendant, Elmer Palmer, murdered his grandfather, Francis Palmer.[148] In his will, Francis left most of his estate to Elmer; fearing Francis might change his will, Elmer preemptively poisoned him.[149] Although Elmer faced a significant prison sentence, New York law at the time did not include an explicit provision stating that Elmer could not inherit his grandfather's estate.[150] Faced with this injustice, the New York Court of Appeals declared that Elmer could not be allowed to benefit through his wrongdoing, and his share of the estate was given to his two aunts, the daughters of the late Francis Palmer.[151] *Riggs v. Palmer* established the general notion that a person must not be enriched through their own wrongdoing, and any

---

(reaffirming the general restitutionary rule that consent is not required to establish duty to pay in emergency cases in which the patient was unable to express her medical need); *see also* BIRKS, *supra* note 137, at 39–40.

143 RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT § 1 (AM. L. INST. 2011) (noting the flexibility of the requirement for the "unjust" enrichment of the defendant); *see* Mark P. Gergen, *What Renders Enrichment Unjust?*, 79 TEX. L. REV. 1927, 1947 (2001). *See generally* Lionel Smith, *Restitution: A New Start?*, *in* THE IMPACT OF EQUITY AND RESTITUTION IN COMMERCE 91 (Peter Devonshire & Rohan Havelock eds., 2018).

144 *See* Hanoch Dagan, *Mistakes*, 79 TEX. L. REV. 1795, 1809–10 (2001); *see also* Ernest J. Weinrib, *Correctively Unjust Enrichment*, *in* PHILOSOPHICAL FOUNDATIONS OF THE LAW OF UNJUST ENRICHMENT 31, 44 (Robert Chambers et al. eds., 2009).

145 *See* RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT § 20 cmt. a (AM. L. INST. 2011).

146 *See id.* § 51(4) (explaining the liability of wrongdoers in unjust enrichment).

147 22 N.E. 188 (N.Y. 1889).

148 *Id.* at 188–89.

149 *Id.*

150 *Id.* at 189.

151 *Id.* at 191.

enrichment generated through a wrong or a crime is to be stripped away from the malfeasor.

In other cases, courts made similar use of the *disgorgement* remedy, used to strip a wrongdoer of any gains obtained through wrongful or harmful activity.[152] The *constructive trust* is an analogous legal instrument.[153] When the defendant unlawfully takes another's asset, the court can construct a legal fiction according to which the defendant is holding the asset as a trustee for the benefit of the true owner.[154] This means that any benefits the defendant made through unlawfully holding the asset are to be given to the original owner in order to prevent unjust enrichment.[155]

Although liability in restitution can exist even without a wrong, any degree of fault by the defendant typically makes the claim stronger, reducing the availability of defenses and allowing augmented remedies. For instance, in a mistaken payment case when the defendant is not a wrongdoer, the defendant-recipient must return any money they received by mistake, but also enjoys robust defense rules. To illustrate, assume a recipient received a large sum of money by mistake but honestly believed the money was a gift from a family member and spent it on an expensive vacation. In such a case, the recipient is considered to have changed their position in good faith in reliance on the payment and is therefore exempt from full restitution.[156] Of course, this type of defense is not available to a wrongdoer, such as a defendant who knowingly took funds that did not belong to them.[157] This defendant can never be considered to have believed, in good faith, that they had a valid legal claim to the money so they cannot enjoy the change of position defense.[158]

Similarly, augmented remedies such as disgorgement of profits or constructive trusts are more commonly available when the defendant is

---

152 Restatement (Third) of Restitution and Unjust Enrichment § 51(Am. L. Inst. 2011) (defining disgorgement as a restitutionary remedy designed to strip a wrongdoer of all ill-gotten gains and characterizing it as typically available in cases where the defendant's intentional wrong enriched her at the expense of another).

153 *See generally* Lionel Smith, *Constructive Trusts and the No-Profit Rule*, 72 Cambridge L.J. 260 (2013).

154 *See id.*

155 *See id.*; Andrew Kull, *Restitution in Bankruptcy: Reclamation and Constructive Trust*, 72 Am. Bankr. L.J. 265, 287 (1998).

156 *See* Restatement (Third) of Restitution and Unjust Enrichment § 65 (Am. L. Inst. 2011).

157 *Id.* at cmt. a; Maytal Gilboa & Yotam Kaplan, *The Costs of Mistakes*, 122 Colum. L. Rev. F. 61, 73 (2022) (explaining the requirements of good faith as an element of the change of position defense).

158 *See* Restatement (Third) of Restitution and Unjust Enrichment § 65 cmt. a (Am. L. Inst. 2011); Gilboa & Kaplan, *supra* note 157, at 73.

a wrongdoer.[159] For instance, a wrongdoer who took another's asset will be liable to return not only the stolen asset, but also any profit illegally obtained through the use of this asset. This form of supracompensatory remedy that is used through a constructive trust is not typically available against a defendant who was enriched through no fault of their own.[160] Thus, if a recipient of a mistaken payment used the sum they received to make a profit, they will usually only be obligated to return the original sum they received and not the profits they obtained by using it, assuming they held and used the sums in good faith.[161]

The rationale behind this basic structure of the law of unjust enrichment is simple. When the defendant is a wrongdoer, a harsher legal response is justified to induce deterrence.[162] As long as wrongdoers can benefit through their wrongs, they have an incentive to act in harmful, wrongful, or illegal ways. To remove such incentive, the law of unjust enrichment operates to strip wrongdoers of their unlawful gains.[163] The more severe the offense, the more important it is to generate deterrence and make sure the harmful behavior is not allowed to be profitable.[164]

The law of unjust enrichment leaves significant room for judicial discretion and creativity.[165] Thus, the defendant's enrichment can be considered "unjust" for many reasons and courts are free to develop this legal category as both new cases and new problems arise.[166] This is a necessary feature of the doctrine as it proves instrumental in the effort to assure deterrence and circumvent opportunism. As we can learn from *Riggs v. Palmer*, wrongdoers will attempt to find loopholes or illegitimate ways to make a profit that are not explicitly forbidden by law;[167] the law of unjust enrichment operates as a safety valve designed to assure such conduct is not allowed to remain profitable.

B. Platform Enrichment

Personalization is key to the business model of social media platforms and to their immense profitability. These platforms charge a

---

159 Restatement (Third) of Restitution and Unjust Enrichment § 51(4) (Am. L. Inst. 2011).

160 *Id.*

161 *See id.* § 51 cmt. a.

162 Grosskopf, *supra* note 19, at 1997–98.

163 *Id.*

164 Deterrence has been recognized as one of the central goals of restitutionary remedies. *See, e.g.*, Restatement (Third) of the Law of Restitution and Unjust Enrichment § 3 cmt. c (Am. L. Inst. 2011) ("Restitution requires full disgorgement of profit by a *conscious wrongdoer*, not just because of the moral judgment implicit in the rule of this section, but because any lesser liability would provide an inadequate incentive to lawful behavior." (emphasis added)).

165 Sherwin, *supra* note 136, at 2107.

166 *Id.* at 2107–08.

167 *See* Riggs v. Palmer, 22 N.E. 188, 188–89 (N.Y. 1889).

premium from advertisers as they are able to specifically target various types of content to users who are likely to find them interesting.[168] Platforms' ability to personalize content makes advertising much more effective and allows advertisers to spend their budgets more efficiently.[169] In the past, advertisers had broad targeting capabilities: advertising beer[170] and cars[171] during the Super Bowl or a bread maker in a housekeeping magazine. Targeted advertising allows advertisers to be much more effective in their advertising by picking much more specific targeting criteria.[172] Platforms enable advertisers to target their tennis shoe ads at people who have actively expressed an interest in tennis, uploaded videos of themselves working out, and have a certain level of income.[173] Data collected along the outgoing vector allows platforms to learn about their users in order to optimize their ability to present users with relevant ads along the incoming vector.[174] From the perspective of the law of restitution, these benefits constitute a form of enrichment.

The data of each individual collected by platforms is worth very little on its own. However, the data of large groups of users are a very valuable resource.[175] In fact, user data is so valuable today that it has even been dubbed "the new oil."[176] Leading platforms have generated substantial gains from user data they collect and analyze.[177] At the same time, individual users are not financially remunerated for the use of their data.[178] Glen Weyl and Eric Posner have used the term

---

168 *See* Marotta et al., *supra* note 51, at 4.

169 *See id.* at 27.

170 *See, e.g.*, Lora Kelley, *Floodgates Open for Beer Ads During Super Bowl*, N.Y. Times (Feb. 10, 2023), https://www.nytimes.com/2023/02/10/business/media/beer-ads-super-bowl.html [https://perma.cc/PSR6-7LWM].

171 *See, e.g.*, Michael Wayland, *Why You Won't See Many Car Ads During Sunday's Super Bowl*, CNBC (Feb. 11, 2024, 5:06 PM), https://www.cnbc.com/2023/02/10/gm-jeep-kia-super-bowl-ads.html [https://perma.cc/A6UJ-W2W7].

172 *See* Leslie K. John, Tami Kim & Kate Barasz, *Ads That Don't Overstep*, Harv. Bus. Rev., Jan.–Feb. 2018, https://hbr.org/2018/01/ads-that-dont-overstep [https://perma.cc/YAJ4-3VS6].

173 *See* Caitlin Dewey, *98 Personal Data Points That Facebook Uses to Target Ads to You*, Wash. Post (Aug. 19, 2016, 10:13 AM), https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/ [https://perma.cc/UQP4-ZF9R].

174 *See* Gordon-Tapiero et al., *supra* note 2, at 647.

175 *See id.* at 647–51.

176 Nisha Talagala, *Data as the New Oil Is Not Enough: Four Principles for Avoiding Data Fires*, Forbes (Mar. 2, 2022, 5:48 PM), https://www.forbes.com/sites/nishatalagala/2022/03/02/data-as-the-new-oil-is-not-enough-four-principles-for-avoiding-data-fires/?sh=3e76a899c208 [https://perma.cc/RA4E-JAYH]; *see The World's Most Valuable Resource is No Longer Oil, but Data*, Economist (May 6, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data [https://perma.cc/6P5V-43LP].

177 *See* Eric A. Posner & E. Glen Weyl, Radical Markets: Uprooting Capitalism and Democracy for a Just Society 231–32 (2018).

178 *See id.*

"technofeudalism"[179] to describe this reality where the value of user data is "distributed to a small number of wealthy savants rather than to the masses."[180]

Not only do users not share in the monetary value of their data,[181] but platforms' problematic personalization processes also generate the harms discussed in Section I.C. These harms are externalized to individual users and to society. This incentive structure allows platforms to continue reaping financial benefits from wrongful, dangerous, and destructive activity. This proposal offers three categories in which platform enrichment can be considered unjust based on problematic personalization processes. The Authors suggest that profits derived from these types of activities should be disgorged. It is important to note that for a gain to be considered *unjust*, a rather high bar must be crossed: not just any gain that makes one feel slightly uncomfortable constitutes *unjust enrichment*. The Authors feel confident, however, that the types of enrichment described here do cross this threshold. The courts will further develop the precise criteria for the unjust enrichment test on a case-by-case basis.

1. Illegal Discrimination

The advertising process on social media platforms allows advertisers to specify a target audience for any ad.[182] Pauline T. Kim and Sharion Scott highlight three mechanisms by which targeting criteria as chosen by the advertiser may generate discriminatory presentation of the ads.[183] First, advertisers can choose their target audience on the platform by specifying personal attributes of users they want to target, as well as attributes of users they want to exclude from seeing their ads.[184] If an advertiser specifies a particular gender or age group as a targeting criterion, the result will be discriminatory.[185] Similarly, if an advertiser decides to exclude people speaking a particular language or of a particular ethnicity, the ad will be presented in a discriminatory fashion.[186]

---

179 *See id.* at 231.

180 *See id.* at 209.

181 *See generally* RadicalxChange Foundation Ltd., *The Data Freedom Act*, RadicalxChange (May 27, 2020), https://www.radicalxchange.org/media/papers/data-freedom-act.pdf [https://perma.cc/4LLC-SXDG].

182 *See* Ali et al., *supra* note 55, at 2.

183 *See* Kim & Scott, *supra* note 39, at 98.

184 *See id.*

185 *See id.*

186 In 2016, *ProPublica* reported that Facebook allowed discriminatory presentation of housing ads. *See* Julia Angwin & Terry Parris Jr., *Facebook Lets Advertisers Exclude Users by Race*, ProPublica (Oct. 28, 2016, 1:00 PM), https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race [https://perma.cc/V24U-UT2S]. Despite Facebook's commitment to preventing discriminatory presentation of such ads in the future, research found that the phenomenon

Second, an advertiser may pick a seemingly neutral targeting criterion that turns out to be highly correlated with a protected attribute and produces a discriminatory outcome.[187] For example, a user's zip code, as well as their membership in ethnic culture groups, are highly correlated with their race.[188] While the correlation between zip code and membership in ethnic culture groups are well established, other attributes may seem innocuous *ex ante*, though analysis of the distribution of their presentation *ex post* may reveal a discriminatory pattern.

Third, advertisers can use what is known as the '"lookalike' audience" tool.[189] This tool allows the advertiser to specify a custom audience and to request that the platform target the ad at users whom the platform determines to be similar to the predefined group.[190] If the sample group defined by the advertiser is biased, the lookalike audience will also be biased.[191]

Targeting criteria as specified by the advertiser, however, are not the only source of bias in advertising. Carefully constructed experiments conducted on Facebook identified discriminatory presentation of ads, even in cases where the criteria specified by the advertiser were

---

persisted. *See* Julia Angwin, *Facebook Says It Will Stop Allowing Some Advertisers to Exclude Users by Race*, ProPublica (Nov. 11, 2016, 10:00 AM), https://www.propublica.org/article/facebook-to-stop-allowing-some-advertisers-to-exclude-users-by-race [https://perma.cc/JY57-FPZS] (highlighting Facebook's commitment to stop discriminatory presentation); *see also* Julia Angwin, Ariana Tobin & Madeleine Varner, *Facebook (Still) Letting Housing Advertisers Exclude Users by Race*, ProPublica (Nov. 21, 2017, 1:23 PM), https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin [https://perma.cc/6XVH-Z4Z2].

187 Kim & Scott, *supra* note 39, at 98; Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau & Alan Mislove, *Potential for Discrimination in Online Targeted Advertising*, 81 Procs. Mach. Learning Rsch. 1, 2 (2018) ("An intentionally malicious—or unintentionally ignorant—advertiser could leverage such data to preferentially target (i.e., include or exclude from targeting) users belonging to certain sensitive social groups (e.g., minority race, religion, or sexual orientation)."). Nondiscrimination law does not consider those who receive unfair treatment in the context of incoming vector personalization, such as "tennis players." *Cf.* Sandra Wachter, *Affinity Profiling and Discrimination by Association in Online Behavioral Advertising*, 35 Berkeley Tech. L.J. 367, 369 (2020) (acknowledging that nondiscrimination law provides protection to certain recognized categories of individuals, but does not take into account "new" categories that may receive unfair treatment).

188 In areas with a high degree of residential segregation, a user's zip code may serve as a proxy for race. *See* Kim & Scott, *supra* note 39, at 98; *see also* Jinyan Zang, *Solving the Problem of Racially Discriminatory Advertising on Facebook*, Brookings Inst. (Oct. 19, 2021), https://www.brookings.edu/research/solving-the-problem-of-racially-discriminatory-advertising-on-facebook/ [https://perma.cc/J8VS-HK6L] (detailing some of the effects of algorithmic discrimination).

189 Kim & Scott, *supra* note 39, at 98.

190 *See* Speicher et al., *supra* note 187, at 11.

191 *See id.* (showing that targeting potential employees based on a "look-alike" audience criterion could also be seen as similar to recruiting via word of mouth).

neutral.[192] For example, despite having the advertiser—in this case the researchers—use the same targeting criteria, ads for cashier positions in supermarkets were presented predominantly to women based on criteria introduced by the platforms in the ad delivery process, as opposed to considerations introduced by the advertiser in the targeting stage.[193] This insight calls for careful consideration of the challenge created by the current structure of the platforms' advertising mechanisms and the financial incentives driving them.[194]

Regardless of how it is generated, discrimination is highly profitable for platforms.[195] Facebook's targeting mechanism enables advertisers to specifically detail the attributes they want to have in their target audience.[196] As explained above, advertisers have a strong interest in targeting their advertisements to people likely to find them relevant and interesting. The fact that platforms allow advertisers to target their ads to users based on data collected about them along the incoming vector makes them an attractive advertising outlet.[197] Thus, platforms are able to charge a higher price for the targeted, discriminatory presentation of such content.

Yet this type of discrimination is illegal. This means that any profits derived from discriminatory ad presentation constitute unjust enrichment. Under section 2000e of Title VII of the Civil Rights Act of 1964,[198] it is illegal to discriminate in the presentation of job ads based on protected attributes such as race, gender, and age.[199] Despite this clear legal standard, leading platforms were found to enable the presentation of job ads in a discriminatory fashion.[200] In 2019, the U.S. Equal Employment

---

192 *See, e.g.*, Ali et al., *supra* note 55, at 19–22; Muhammad Ali, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove & Aaron Rieke, *Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging*, 2021 Web Search and Data Mining 13, 20 ("Our findings suggest that Facebook is wielding significant power over political discourse through its ad delivery algorithms . . . .").

193 *See* Ali et al., *supra* note 55, at 21.

194 *See id.* at 24–25 (calling to consider the policy implications of the study's findings).

195 *See* Viljoen, *supra* note 17, at 588 (Google reported 134.81 billion dollars in advertising revenue in 2019); *Advertising Revenues Generated by Facebook Worldwide from 2017 to 2027*, Statista (Aug. 2023), https://www.statista.com/statistics/544001/facebooks-advertising-revenue-worldwide-usa/ [https://perma.cc/69JQ-PWK7] (Facebook generated 113.64 billion dollars in advertising revenues worldwide in 2022); Jacqueline Zote, *Instagram Statistics You Need to Know for 2023*, Sprout Soc. (Mar. 6, 2023), https://sproutsocial.com/insights/instagram-stats/ [https://perma.cc/ST3A-UMJV] (Instagram made 43.2 billion dollars on advertisements in 2022).

196 *See* Zang, *supra* note 188.

197 *See id.*

198 42 U.S.C. § 2000e.

199 The Civil Rights Act of 1964 §§ 703–716, 42 U.S.C. § 2000e.

200 *See, e.g.*, Basileal Imana, Aleksandra Korolova & John Heidemann, *Auditing for Discrimination in Algorithms Delivering Job Ads*, 2021 The Web Conf. 3767, 3769 (demonstrating that presentation of ads on Facebook and LinkedIn can be skewed by gender); *see also* Alexia Fernández Campbell, *Job Ads on Facebook Discriminated Against Women and Older Workers, EEOC Says*, Vox (Sept. 25, 2019, 2:20 PM), https://www.vox.com/identities/2019/9/25/20883446/

Opportunity Commission found that ads presented on Facebook discriminated against women and older workers.[201] Numerous other platforms have also been found to present housing and employment ads in a discriminatory manner.[202] As explained below, this proposal to treat gains from discriminatory platform advertising as unjust enrichment is necessary to remove the profitability of this practice and platforms' incentive from participating in it. It is, however, not meant to replace any existing regulatory mechanism designed to combat discrimination, but rather is meant as an additional tool in the legal antidiscriminatory arsenal.

2. The Abuse of Vulnerable Users

This Section argues that platforms' profits must be considered unjust enrichment when they originate from predatory practices that target vulnerable users and attempt to monetize their vulnerability. Children and teens are among the most vulnerable groups of users of social media. They are at an age when they are "less privy to marketing techniques and so more susceptible to the tactics of online marketers and their deceptive trade practices."[203] Thus, they "may be deceived by an image or a message that likely would not deceive an adult."[204] This explanation reflects a recognition that at times content presented to an adult may not be troubling or cause harm, but that the very same

---

facebook-job-ads-discrimination [https://perma.cc/827S-JFYH] (finding that Facebook presented ads in a way that discriminated against women and older users); Anja Lambrecht & Catherine Tucker, *Apparent Algorithmic Discrimination and Real-Time Algorithmic Learning in Digital Search Advertising* (Apr. 15, 2021) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3570076 [https://perma.cc/8UKY-36HV] (finding that Google presented ads for disadvantageous jobs to users who had previously searched for Black names compared to the jobs advertised to users who had previously searched for White names).

201 Press Release, ACLU, In Historic Decision on Digital Bias, EEOC Finds Employers Violated Federal Law when They Excluded Women and Older Workers from Facebook Job Ads (Sept. 25, 2019), https://www.aclu.org/press-releases/historic-decision-digital-bias-eeoc-finds-employers-violated-federal-law-when-they [https://perma.cc/74QM-MFFD] (reporting on the decision); *Letters of Discrimination*, U.S. Equal Emp. Opportunity Comm'n (July 5, 2019), https://www.onlineagediscrimination.com/sites/default/files/documents/eeoc-determinations.pdf [https://perma.cc/4LNE-F3N5].

202 *See* Ali et al., *supra* note 55, at 20–22 (observing significant skews in the presentation of ads for housing and employment along gender and racial lines); Imana et al., *supra* note 200, at 3774–75 (demonstrating that presentation of ads on Facebook and LinkedIn can be skewed by gender).

203 Shannon Finnegan, *How Facebook Beat the Children's Online Privacy Protection Act: A Look into the Continued Ineffectiveness of COPPA and How to Hold Social Media Sites Accountable in the Future*, 50 Seton Hall L. Rev. 827, 829 (2020).

204 J. Howard Beales, III, *Advertising to Kids and the FTC: A Regulatory Retrospective That Advises the Present*, 12 Geo. Mason L. Rev. 873, 873 (2003).

content shown to a child may be harmful.[205] The Federal Trade Commission has been quite successful in enforcing the prohibition on deceptive practices in the context of advertising targeted at children.[206] On social media, many of the products or services being promoted do not appear in the form of an advertisement. Instead, products are promoted through what is known as "stealth advertising"—the presentation of seemingly organic content by influencers.[207] Such mechanisms have been used, for example, to circumvent rules limiting the advertising of cigarettes to teens.[208] Instead of targeting a potential audience through regular ads, influencers are now paid to present content promoting cigarette use, enabling them to avoid direct application of advertising restrictions.

Another way that social media platforms unjustly enrich themselves involves the viral spread of "challenges" among younger crowds. This genre of content has spread on platforms such as TikTok, Instagram, and YouTube and often includes children and teens participating in dangerous activities and self-harm.[209] Famous challenges include the

---

205 Raffaello Rossi & Agnes Nairn, *How Children Are Being Targeted with Hidden Ads on Social Media*, Conversation (Nov. 3, 2021, 8:24 AM), https://theconversation.com/how-children-are-being-targeted-with-hidden-ads-on-social-media-170502 [https://perma.cc/VF8K-6THY] (acknowledging the vulnerability of children); Raffaello Rossi & Agnes Nairn, What Are the Odds? The Appeal of Gambling Adverts to Children and Young Persons on Twitter 4 (2021), https://www.bristol.ac.uk/media-library/sites/management/documents/what-are-the-odds-rossi-nairn-2021.pdf [https://perma.cc/MPK8-ZXEG] (finding gambling advertisements are far more appealing to children and young people than adults).

206 *See In re* Lewis Galoob Toys, Inc., No. C-3324, 114 F.T.C. 187, 214–15 (1991) (settled by consent order); *In re* Hasbro, Inc., No. C-3447, 116 F.T.C. 657, 667 (1993) (settled by consent order); *see also In re* Mattel, Inc., No. C-2071, 79 F.T.C. 667, 671–72 (1971) (settled by consent order). Along the outgoing vector, the collection of data from children is restricted by the Children's Online Privacy Protection Act. 15 U.S.C. §§ 6501–6506 (2018); 16 C.F.R. § 312.9 (2019).

207 *See* Rossi & Nairn, *supra* note 205 (detailing how stealth advertising is particularly dangerous when targeted at children).

208 Megan Cerullo, *Health Groups Call Out Tobacco Marketing Aimed at Teens on Social Media*, CBS News (May 22, 2019, 7:00 PM), https://www.cbsnews.com/news/health-groups-call-out-tobacco-and-e-cigarette-marketing-aimed-at-teens-on-social-media/ [https://perma.cc/P5WQ-ECFG]; *see The Effect of Social Media Ads on Teen Behavior*, Stop Med. Abuse (Mar. 29, 2018), https://stopmedicineabuse.org/blog/details/the-effect-of-social-media-ads-on-teen-behavior/ [https://perma.cc/XMP9-QM6F]; Lisa Rapaport, *Click Bait Ads Are Tied to Teen Smoking*, Reuters (Jan. 3, 2018, 3:57 PM), https://www.reuters.com/article/us-health-teens-tobacco-ads/click-bait-ads-are-tied-to-teen-smoking-idUSKBN1ES1XH [https://perma.cc/F9HC-3F72]; *see also Just How Harmful Is Social Media? Our Experts Weigh-In*, Columb. Mailman Sch. Pub. Health (Sept. 27, 2021), https://www.publichealth.columbia.edu/public-health-now/news/just-how-harmful-social-media-our-experts-weigh [https://perma.cc/QW3G-WRP6] (describing the dangers of social media more generally).

209 *See* J. Ortega-Baron, J.M. Machimbarrena, I. Montiel & J. González-Cabrera, *Viral Internet Challenges Scale in Preadolescents: An Exploratory Study*, 42 Current Psych. 12530 (2023).

tide pod challenge,[210] the blackout challenge,[211] cinnamon challenge,[212] blue whale challenge,[213] ice and salt challenge,[214] and many more equally extremely dangerous challenges.[215] TikTok's algorithm promotes videos with trending hashtags which enables the poster to get more views and engagement as these videos will be pushed higher up in users' feeds.[216] The virality of such challenges is, of course, highly profitable for platforms by increasing user engagement.[217] This would explain platforms' support and promotion of such content. The Authors argue that the clear harmfulness of such activities must mean the enrichment that follows should be considered unjust and be stripped away from platforms. Anything less will maintain the profitability of such practices for platforms and thus perpetuate these harmful occurrences.

---

210 Lindsey Bever, *Teens Are Daring Each Other to Eat Tide Pods. We Don't Need to Tell You That's a Bad Idea*, WASH. POST (Jan. 17, 2018, 8:07 PM), https://www.washingtonpost.com/news/to-your-health/wp/2018/01/13/teens-are-daring-each-other-to-eat-tide-pods-we-dont-need-to-tell-you-thats-a-bad-idea/ [https://perma.cc/UJF5-7CTY]; Claire McCarthy, *Why Teenagers Eat Tide Pods*, HARV. HEALTH PUBL'G (Jan. 30, 2018), https://www.health.harvard.edu/blog/why-teenagers-eat-tide-pods-2018013013241 [https://perma.cc/7MLY-RHPP].

211 Seren Morris, *10-Year-Old Girl Dies in 'Blackout Challenge' Circulating on TikTok*, NEWSWEEK (Jan. 22, 2021, 11:36 AM), https://www.newsweek.com/girl-dies-blackout-challenge-circulating-tiktok-1563705 [https://perma.cc/QSZ5-ZEDS] (describing the challenge, "which encourages people to try and pass out by restricting their airflow").

212 *"Cinnamon Challenge" Dangerous to Lungs, New Report Warns*, CBS NEWS (Apr. 22, 2013, 12:49 PM), https://www.cbsnews.com/news/cinnamon-challenge-dangerous-to-lungs-new-report-warns/ [https://perma.cc/8BXR-2BKJ] (providing that the cinnamon challenge involved trying to swallow a spoonful of cinnamon within sixty seconds).

213 Ortega-Baron et al., *supra* note 209, at 12531 (describing the Blue Whale Challenge, "which consists of a chain of challenges that contain self-harming acts that lead up to the person's suicide"); *see also* Mahesh Mahadevaiah & Raghavendra B. Nayak, *Blue Whale Challenge: Perceptions of First Responders in Medical Profession*, 40 INDIAN J. PSCYH. MED. 178, 179 (2018).

214 Forrest Saunders, *'Salt and Ice Challenge' Leaves Iowa Kids with Severe Burns*, KCRG (Jan. 25, 2019, 12:03 AM), https://www.kcrg.com/content/news/Salt-and-ice-challenge-leaves-Iowans-with-severe-burns--504847271.html [https://perma.cc/RR8Q-VPSP] (detailing the salt and ice challenge, which involved putting table salt on the skin and then pressing ice into it, gave teens second- and third-degree burns).

215 *See Dangerous Social Media Challenges: Understanding Their Appeal to Kids*, HEATHY CHILDREN (Sept. 11, 2023), https://www.healthychildren.org/English/family-life/Media/Pages/Dangerous-Internet-Challenges.aspx [https://perma.cc/4VFA-2EVL].

216 Jami Reetz, *TikTok Trends: How to Find Them and Make Them Your Own*, BAZAAR VOICE (Nov. 6, 2023), https://www.bazaarvoice.com/blog/tiktok-trends-how-to/ [https://perma.cc/PX9T-GXBA]; Christina Newberry, *The TikTok Algorithm Explained + Tips to Go Viral*, HOOTSUITE: BLOG (Feb. 8, 2023), https://blog.hootsuite.com/tiktok-algorithm/ [https://perma.cc/HU86-JWEQ]; Marcus Johnson & Aran Sonnad-Joshi, *Are Social Media Challenges a Force for Good or Evil?*, S. ONLINE (Oct. 20, 2021), https://thesoutherneronline.com/85043/front-slideshow/are-social-media-challenges-a-force-for-good-or-evil/ [https://perma.cc/36LJ-MHMH]; Katie Elson Anderson, *Getting Acquainted with Social Networks and Apps: It's Time to Talk About TikTok*, 37 LIBR. HI TECH NEWS 7, 9 (2020) (describing that TikTok users who access the "Discover" icon on the platform's interface will be presented with current trending hashtags).

217 *See supra* Section I.B.

The current harmful dynamic and the unwillingness of platforms to act decisively can be observed in the self-regulation of challenges by platform. Thus, when a certain trend causes what is seen to be "too much" damage or generates "too much" negative publicity for the platform, TikTok has taken action.[218] Yet, before a trend reaches such extremity, it is allowed to continue undisturbed. Despite claims by platforms such as YouTube and TikTok that they would act against the spread of such challenges,[219] the first half of 2022 saw their continued spread.[220] Kate Tilleczek has called attention to the revenue generated by TikTok from the spread of such content, saying, "You leave [regulation] in the hands of folks who are making billions of dollars to do the right thing by kids, and I'm always thinking: 'They're not going to do that.'"[221] Tillaczek echoes similar sentiments expressed by Frances Haugen: expecting platforms to hurt their revenue by limiting the enrichment they generate from wrongful practices is like leaving the cat to guard the cream.[222] A legal doctrine that prevents platforms from wrongfully becoming enriched at the expense of vulnerable groups is thus necessary to bring about a real change in the way platforms design their algorithms.

Another vulnerable group suffering due to the use of social media includes people who were the subject of human trafficking.[223] The

---

218 For example, when TikTok determined that a viral hashtag promoted a challenge that was deemed dangerous, it removed the hashtag promoting the challenges, lowering its spread. *See* Michael Ordoña, *TikTok Bans Milk Crate Challenge (Because It's Super-Dangerous to Fall off Crates?)*, L.A. Times (Aug. 27, 2021, 3:43 PM), https://www.latimes.com/entertainment-arts/story/2021-08-27/tiktok-bans-milk-crate-challenge [https://perma.cc/2P5F-366E]. In another case, TikTok accompanied videos of a dangerous challenge with a warning. *See* Jamie Harris, *Ticked Off TikTok Will Now Warn Teens About Dangerous Viral Challenges They're Searching*, The Sun (Feb. 23, 2022, 12:13 PM), https://www.thesun.co.uk/tech/17735154/tiktok-warn-teens-dangerous-viral-challenges/ [https://perma.cc/HE6X-4CFA].

219 *See* McCarthy, *supra* note 210; Michelle Toh, *Tide Pod Challenge: YouTube Is Removing 'Dangerous' Videos*, CNN Bus. (Jan. 18, 2018, 8:25 AM), https://money.cnn.com/2018/01/18/technology/tide-pod-challenge-video-youtube-facebook/ [https://perma.cc/B2JQ-RCX4]; *TikTok Says It's Cracking Down on Dangerous Challenges. Will It Be Enough?*, CBC Kids News (Nov. 18, 2021) [hereinafter CBC Kids News], https://www.cbc.ca/kidsnews/post/tiktok-announces-plan-to-address-dangerous-challenges-and-hoaxes [https://perma.cc/TUF8-5465].

220 *See* Rebecca Rhodes, *The Worst (and Most Dangerous) TikTok Challenges of June 2022*, Know Your Meme (June 14, 2022), https://knowyourmeme.com/editorials/meme-insider/the-worst-and-most-dangerous-tiktok-challenges-of-june-2022 [https://perma.cc/QK6Q-LQZJ].

221 *See* CBC Kids News, *supra* note 219.

222 *See Hearing*, *supra* note 9.

223 As early as 2017, reports surfaced of ads on Facebook being used to recruit members for a Mexican drug cartel. *See* Zorayda Gallegos, *Mexico's Jalisco Drug Cartel Uses Facebook to Recruit New Hitmen*, El Pais (Aug. 3, 2017, 4:39 PM), https://english.elpais.com/elpais/2017/08/01/inenglish/1501585590_499112.html [https://perma.cc/E3S6-ZXQR]; *see also*, Scheck et al., *supra* note 1. For an in-depth discussion of how human traffickers use social media, see Nicola A. Boothe, *Traffickers' "F"ing Behavior During a Pandemic: Why Pandemic Online Behavior Has Heightened*

Facebook Files revealed not only that human traffickers widely used Facebook for trafficking purposes, but also that Facebook was aware of this practice dating back to at least 2018.[224] It was only once Apple threatened to remove Facebook and Instagram from the App Store that the platform took action and removed troubling content.[225]

To ensure that platforms' incentives are more aligned with the public interests, they must be stripped of wrongful gains generated from harmful personalization using the law of unjust enrichment.

3. Socially Harmful Personalization

Finally, this Article suggests that platform profits can additionally be considered unjust enrichment when personalization is connected with the promotion of socially harmful content. At this point in time, several years after implementing the downstream MSI optimization metric, platforms are aware of the type of content their personalization algorithm promotes. They are knowingly and actively promoting harmful, divisive, and extreme content that contributes to extremism, polarization, and democratic erosion. Platforms utilize problematic personalization techniques as it enables them to increase the time users spend on the platform as well as users' interaction with the platform.

This type of enrichment should be considered unjust, taking into account the broad societal harms caused by problematic personalization algorithms. The spread of disinformation promotes distrust and blurs users' ability to differentiate between what is true and what is false. Pushing users to polarizing extremes is harmful as it creates an ever-increasing divide between users with different starting points. It pushes people to adopt extreme positions and dangerous conspiracy theories. Platforms' choices regarding what content to present to users causes them to become locked into filter bubbles that preclude meaningful discourse, which is central to a functioning and flourishing democracy. Users locked into an echo chamber surrounded by people reinforcing their positions and pushing them to further extremes—no longer being able to differentiate between reality and conspiracy—may be pushed to take extreme, violent actions offline.

Moreover, platforms' problematic personalization structure allowed for the promotion of an organized disinformation campaign by foreign governments in the months leading up to the 2016 U.S. presidential election. The Russian government in particular used social media

---

the Urgency to Prevent Traffickers from Finding, Friending and Facilitating the Exploitation of Youth via Social Media, 22 GEO. J. GENDER. L. 533 (2021).

224 *See* Scheck et al., *supra* note 1.

225 Clare Duffy, *Facebook Has Known It Has a Human Trafficking Problem for Years. It Still Hasn't Fully Fixed It*, CNN BUS. (Oct. 25, 2021, 7:33 AM) https://edition.cnn.com/2021/10/25/tech/facebook-instagram-app-store-ban-human-trafficking/index.html [https://perma.cc/US3Z-E7K9].

during this period to attempt to manipulate the outcome of the election. They operated thousands of fake profiles for purposes of "sowing discord in the US political system,"[226] while targeting individuals who would be most susceptible to their messages.[227]

The fact that bad actors are able to spread their harmful, manipulative content so effectively to users who are likely to be susceptible to it is based on platforms' algorithms and their optimization metrics. The algorithms actively promote such harmful content, suggesting that susceptible users join groups, like pages, or follow trending hashtags promoting such content. By structuring their algorithms in such a way, platforms are actively and knowingly becoming unjustly enriched at the expense of society and their users.

III. Comparative Advantages & Implications

This Part offers a discussion of the proposal as outlined in Part II. It explains the rationale of using the law of unjust enrichment in the case of wrongful gains generated by platforms' harmful personalization and details the advantages of this proposal. This Part also offers prediction of platforms' possible responses to the implementation of this proposal.

A. The Comparative Advantages of Unjust Enrichment Law

The challenges highlighted in Section I.C are well known and widely researched. They have been recognized for several years as a harmful byproduct of the way platforms personalize content for their users.[228] Much thought has been given to overcoming them. This Article does not argue that the law of unjust enrichment is the only way to contend with these issues or that other legal routes should be abandoned. The Authors do posit, however, that addressing these harms through the lens of unjust enrichment offers significant advantages that other tools do not and seems appropriate for several reasons.

---

226 1 Robert S. Mueller III, Report on the Investigation Into Russian Interference in the 2016 Presidential Election 14 (2019) ("The IRA conducted social media operations targeted at large U.S. audiences with the goal of sowing discord in the U.S. political system.").

227 *Id.* at 19 ("The IRA's U.S. operations sought to influence public opinion through online media and forums."). *See generally* S. Select Comm. on Intel., 116th Cong., Rep. on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election (2020).

228 *See, e.g.*, Ashley Smith-Roberts, *Facebook, Fake News, and the First Amendment*, 95 Denv. L. Rev. F. 118, 119 (2018); Zeynep Tufekci, *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 Colo. Tech. L.J. 203, 203 (2015); Christopher A. Bail et al., *Exposure to Opposing Views on Social Media Can Increase Political Polarization*, 115 Proc. Nat'l Acad. Scis. 9216, 9216 (2018); Daniel Susser, Beate Roessler & Helen Nissenbaum, *Technology, Autonomy, and Manipulation*, 8 Internet Pol'y Rev. 1, 1 (2019).

1. Harms Versus Gains

The law of unjust enrichment generally focuses on gains rather than on harms.[229] This can offer several advantages. The first advantage relates to deterrence. Through the disgorgement remedy, courts can strip wrongdoers of any ill-gotten gains;[230] such measure of recovery is often necessary to assure that the wrongful activity does not remain profitable and therefore does not persist.[231] Personalization is an immense profit engine for social media platforms.[232] The main income source of these platforms stems from selling personalized advertising services.[233] In the case of Facebook, alarming percentages of these astronomical profits come from fake news.[234] As long as platforms are allowed to benefit through abusing personalization technologies, they will continue to do so, and the harms of malevolent personalization, as described above, will persist. The most effective way to appropriately deter platforms and assure that such activities cease is to strip them of any gains obtained through harmful personalization tactics.

The second advantage of using gains-based recovery pertains to situations in which harms are difficult to measure, but gains can be more easily identified and quantified.[235] Harms such as political polarization and democratic erosion are real and horrifying, but they are probably too abstract and spread over too many unidentified victims to be a basis for a tort, harm-based, claim. To establish such a claim, some identified victim of harm must prove the magnitude of harm caused to them.[236] Even if this identification would be possible in some cases, difficulties in proving harms, their magnitudes, and the identity of victims would make suits prohibitively costly, thus crippling their deterrent effect.

---

229 *See* Laycock, *supra* note 134, at 1283 (observing the essential differences that distinguish restitution from compensation); Maytal Gilboa & Yotam Kaplan, *The Mistake About Mistakes: Rethinking Partial and Full Restitution*, 26 Geo. Mason L. Rev. 427, 427–28 (2018).

230 Restatement (Third) of Restitution and Unjust Enrichment § 51(4) (Am. L. Inst. 2011).

231 Grosskopf, *supra* note 19, at 1997–98.

232 *See* Viljoen, *supra* note 17, at 588–89 ("In 2019, Google reported $134.81 billion in advertising revenue out of $160.74 billion in total revenue. In the first quarter of 2020, Facebook's total advertising revenue amounted to $17.44 billion, compared to $297 million in revenue from other streams.").

233 *See id.*

234 *See* Cohan, *supra* note 121.

235 For an explanation of the prevalence of gains-based remedies in contract law, see Steve Thel & Peter Siegelman, *You* Do *Have to Keep Your Promises: A Disgorgement Theory of Contract Remedies*, 52 Wm. & Mary L. Rev. 1181, 1181–82 (2011).

236 *See* Maytal Gilboa & Yotam Kaplan, *Loser Takes All: Multiple Claimants & Probabilistic Restitution*, 10 U.C. Irvine L. Rev. 907, 911 (2020).

Conversely, the law of unjust enrichment focuses on the behavior of the actor causing the harms and on the enrichment generated by it.[237] The unjust enrichment doctrine does not require identifying an individual harmed by the activity.[238] Most important, as the monetary focus is on the profit generated, it does not require quantifying the harms caused. Of course, this is not to be taken to mean that measuring unjust platform profits in cases of unfair personalization is costless or even easy. But it is possible and well within the reach of routine practices of civil litigation.

2. Calculating Gains

Calculating compensation for damages is never an easy task. It requires asking what would have happened if the harmful action had not occurred, thus assessing the value of an alternative sequence of events. Despite the complexity, courts conduct such calculations on a regular basis in a variety of civil proceedings. In many cases, calculating gains can be significantly easier. For example, in cases of democratic erosion or political polarization, harms are spread among large segments of the population or suffered by society as a whole and are practically impossible to measure and quantify in monetary terms. By comparison, gains in such cases are much easier to assess and are monetary in nature. The relevant gains are accumulated by just one relevant and easily identifiable party: the gains of a specific platform generated by a particular type of activity—that is, the harmful personalization. Although the public does not have access to detailed accounts of platforms' revenue, platforms do indeed hold information regarding the revenue they made from the personalization of different types of content to various audiences.[239] Courts can mandate the disclosure of such documentation as necessary for a precise calculation of platforms gains.

In some cases, calculating gains is quite straightforward. Ads that target vulnerable groups and present them with harmful content allow platforms to become unjustly enriched at the expense of members of these vulnerable groups. This was the case regarding advertisements promoting human trafficking on Facebook. As detailed in Section II.B.2, Facebook knew it was being used to facilitate human trafficking.

---

237 See RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT § 3 cmts. a–c (AM. L. INST. 2011) (explaining that disgorgement of profits can be granted even when plaintiff did not prove any loss); Gilboa & Kaplan, *supra* note 236, at 911 (explaining that unjust enrichment claims focus on the enrichment by the defendant, who is relatively easier to identify compared to the plaintiff).

238 *E.g.*, RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT § 1 cmt. a (AM. L. INST. 2011) (explaining that the formula "at the expense of another" does not require plaintiffs to show that they have suffered loss but rather to focus on the defendants' benefit instead).

239 *See, e.g.*, Duffy, *supra* note 225.

Despite Facebook's commitment to fight human trafficking on its platforms, a Facebook report found that the company sold over $150,000 worth of advertisements facilitating the sale and sexual exploitation of victims of human trafficking.[240] This is a clear example in which the platform has become unjustly enriched and this enrichment was indeed measured by the platform itself. By comparison, the precise harms suffered by human trafficking victims as a result of this campaign are much more difficult to identify and measure.

Another category of advertisements that allows platforms to become unjustly enriched include profiles abusing the social media platforms in order to manipulate the public, purposely promote disinformation, and undermine democratic processes. For example, in the period leading up to the 2016 U.S. presidential election, the Russian Government posted paid ads on Facebook with the goal of sowing discord and mistrust within the American public.[241] In 2018, Congress released over 3,500 such ads.[242] Income generated from the publication of such ads can be considered unjust enrichment generated at the expense of society and should be disgorged. In June 2022, Facebook identified and removed profiles engaged in coordinated inauthentic behavior ("CIB"). CIB is when "groups of pages or people work together to mislead others about who they are or what they're doing."[243] The ads presented to such profiles generated income for the platform while allowing the continued activity of profiles engaged in CIB to the detriment of society. Revenue generated by advertisements presented to profiles later removed due to CIB should be viewed as unjust enrichment generated at the expense of society. Research conducted by *Wired* found that Facebook's parent company, Meta, made at least $30.3 million between July 2018 and April 2022 from advertisements posted by profiles which were later removed from the platforms due to CIB.[244] Again, the harms of Facebook activity in such a case are impossible to estimate, but the gains are easily calculated by researchers.

---

240 *See id.*

241 Kurt Wagner, *Congress Just Published All the Russian Facebook Ads Used to Try and Influence the 2016 Election*, Vox (May 10, 2018, 12:48 PM), https://www.vox.com/2018/5/10/17339864/congress-russia-advertisements-facebook-donald-trump-president [https://perma.cc/J4PU-MHFB].

242 *Social Media Advertisements*, U.S. House of Representatives Permanent Select Comm. on Intel., https://democrats-intelligence.house.gov/social-media-content/social-media-advertisements.htm [https://perma.cc/SG3C-J7FL] (including links to all advertisements posted by the Russian government in the period leading up to the 2016 U.S. Presidential election).

243 Nathanial Gleicher, *Coordinated Inauthentic Behavior Explained*, Meta (Dec. 6, 2018), https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/ [https://perma.cc/Q24R-FGX2].

244 Vittoria Elliott, *Meta Made Millions in Ads from Networks of Fake Accounts*, Wired (June 23, 2022, 7:00 AM), https://www.wired.com/story/meta-is-making-millions-from-fake-accounts/ [https://perma.cc/PFQ5-6Z6P].

When discussing the promotion of increasingly extreme content on social media, YouTube's recommendation algorithm is often identified as "one of the greatest engines of extremism."[245] Often before users can watch a video on YouTube, they must watch the ad presented before the video. Ads for companies like Adidas, Amazon, and Hershey were presented to users on YouTube before videos promoting extreme content.[246] Ads presented before an extreme, polarizing video generate gains for platforms while causing harm to society. Since these ads can easily be connected to the precise type of content presented to a user, the calculation of the platform's gains is a relatively simple challenge.

As discussed in Section I.A, platforms promote content to users in order to increase the time they spend interacting with the platform. Not only does the increased engagement allow platforms to learn more about their users, but during the increased time spent on the platform, users can be presented with more ads. When the personalized content is harmful—whether because of the content itself or because of the nature of the audience it is presented to—the added revenue that platforms make from their extended ability to present more ads should be viewed as unjust enrichment, generated at the expense of vulnerable users and of society at large. While the calculation of this type of enrichment is more complex than simply adding numbers of payments made by advertisers, it is not outside the scope of calculations that courts conduct on a daily basis.

3. Rules Versus Standards

The basic maxim of the law of unjust enrichment provides a flexible standard rather than a clear-cut rule.[247] The distinction between rules and standards is central to legal design.[248] Rules provide sharp dichotomies between two legal categories and leave little room for discretion.[249] Standards on the other hand provide a fuzzier distinction, allow for more discretion in their application, and are more sensitive to context and to the specific detail of each case.[250] Naturally, a mature legal system

---

245 Rabbit Hole, *supra* note 57, at 16:00; *see also id.* at 16:57 ("YouTube was, essentially, built to pull people into these polarizing rabbit holes . . . it's happening not by accident but by design."); Tufekci, *supra* note 122.

246 *See* Rabbit Hole, *supra* note 57, at 15:38 ("CNN reports that YouTube ran ads from large brands like Adidas, Amazon, and Hershey before videos which promoted extreme content.").

247 *See* Sherwin, *supra* note 136, at 2086–87.

248 For an analysis of this distinction, see Duncan Kennedy, *Form and Substance in Private Law Adjudication*, 89 Harv. L. Rev. 1685 (1976); Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 Duke L.J. 557 (1992).

249 Kennedy, *supra* note 248, at 1685.

250 *Id.*

utilizes both rules and standards as each form of legal norm offers other types of advantages.

In the present context, the law of unjust enrichment offers a flexible standard under which cases of harmful personalization can be decided. At this stage it is not appropriate to have a central regulator offer a single clear-cut distinction that would determine when platform personalization constitutes unjust enrichment. Any such determination would be arbitrary and insufficiently sensitive to context and detail. Rather, the Authors see it as preferable to leave those determinations to the discretion of the courts in specific cases, thus allowing the distinction to naturally develop over time while incorporating information from various cases. As demonstrated above, it is easier to say, in *specific instances*, that platform personalization has been unjust. Rather than try to generalize such cases into a strict rule at this stage, it would be more prudent to display patience and allow the law to develop organically through the courts.

Several attempts have been made to regulate the personalization process on platforms.[251] Attention has been focused on limiting the spread of disinformation—with particular attention on health disinformation—especially in the context of the harms disinformation has generated to democracy and democratic institutions.[252] Attempts have been made to limit platforms' ability to manipulate users by presenting them with personalized content, for example, in the context of experimentation.[253] Another type of regulatory attempt to overcome harms of platforms' personalization appears in the form of increased transparency requirements.[254] While the Authors commend such regulatory attempts, these attempts have their inherent limitations. Social media platforms operate in a highly innovative and rapidly changing environment. The slow and cumbersome process of regulation is ill fitted to address the challenges created by the rapid changes and dynamic character of innovative developments.[255] Information gaps introduce similar difficulties. While innovators are typically well acquainted with their innovation and the market conditions in which they operate, regulators' acquaintance often lags behind.[256] This can make it hard for regulators to understand how their decisions will impact the innovative practice and the market

---

251 *See* Justice Against Malicious Algorithms Act, H.R. 5596, 117th Cong. (2021); Health Misinformation Act, S. 2448, 117th Cong. (2021); Algorithmic Justice and Online Platform Transparency Act, S. 1896, 117th Cong. (2021).

252 *See, e.g.*, S. 2448.

253 *See* Deceptive Experiences To Online Users Reduction (DETOUR) Act, S. 1084, 116th Cong. (2019); S. 2448.

254 *See* S. 1896.

255 *See* Sofia Ranchordás, *Innovation-Friendly Regulation: The Sunset of Regulation, the Sunrise of Innovation*, 55 JURIMETRICS J. 201, 206 (2015).

256 *See id.* at 203.

it operates within.[257] The doctrine of unjust enrichment, as a common law doctrine, can be applied by courts to the details of a particular case brought before them. Thus, the court can focus on the unjust enrichment in a particular case using information brought forth by relevant plaintiffs without the need to examine the overall functioning of platforms. This allows flexibility to decide on the merits of a particular lawsuit and allows courts to tailor the response to the facts of a particular case.

The other advantage of a court applied common law doctrine, as opposed to regulation, has to do with the rigid nature of regulation as opposed to the flexible nature of platform personalization. If regulation would prohibit a certain type of personalization or limit the use of a particular optimization metric, platforms could find a way to tweak their activity to ensure it was no longer covered by the regulation. Under the doctrine of unjust enrichment, a court would examine the platforms' behavior. If it finds that the personalization process has been unjust, it can then disgorge any profits generated by it regardless of the tools that the platforms used for their unjust behavior. Applying the doctrine of unjust enrichment to platform personalization does not require platforms to behave in a particular way. Instead, it seeks to disincentivize them from using optimization metrics or personalization algorithms that allow them to become unjustly enriched at the expense of society. Application of this doctrine shifts the responsibility back to platforms and allows them to pick any personalization process and metrics as long as they do not unjustly harm society.

4. The Diversity of Plaintiffs

A key advantage of the use of unjust enrichment doctrine is that claims in unjust enrichment can be brought to the courts by various types of plaintiffs. This can assist in avoiding regulatory capture[258] and in utilizing comparative informational and institutional advantages of diverse potential plaintiffs.[259]

First, a claim in unjust enrichment can be brought to court by a plaintiff at whose expense the defendant was unjustly enriched. In the case of platform enrichment, any user will probably satisfy these

---

257 *See id.*

258 *See* Glover, *supra* note 26, at 1154.

259 *Id.* ("Moreover, public civil enforcers in some regulatory areas suffer informational disadvantages. Those disadvantages arise for a simple reason: the best sources of information about private wrongs are often the parties themselves, because they tend to have superior knowledge regarding the costs and benefits of given activities, the costs of reducing risks of harm, and the probability or severity of risk." (footnote omitted)); *see also* Steven Shavell, *Liability for Harm Versus Regulation of Safety*, 13 J. LEGAL STUD. 357, 359–65 (1984) (highlighting informational advantages of private versus public regulation).

requirements as the platform benefits by misusing its users' data.[260] Naturally, the platform's enrichment at the expense of a specific individual user is marginal; platform profits come from the aggregation of the data of hundreds of millions of users.[261] To bridge this gap, some form of aggregated legal claim akin to a class action[262] will have to be used to strip the platform of the full amount of its ill-obtained gains. Under such a scheme, the actual plaintiff, representing the group of users, will receive a part of any monetary reward granted by the court at the end of the proceedings.[263] This reward, as in a typical class action scenario, is meant to encourage the group representative to bring the claim to the court, acting as a "private attorney general" and promoting the overall social interest.[264] This incentive is beneficial in recruiting individual plaintiffs to act for the greater good; this is advantageous in the common instances in which such individuals enjoy informational advantages over central regulators.[265] It will often be the case, for instance, when online "challenges" trending among adolescents cause personal injury; private plaintiffs more easily obtain information in such cases than central regulators.[266]

The share of the award going to the representative plaintiff will usually remain relatively small. The court will divide the lion's share of any award equally among platform users. This just outcome not only deters platforms from abusing their power, but it also makes intuitive sense concerning the implicit bargain between the parties. Social media platforms provide services free of charge with users effectively paying by allowing the platforms to mine and use their data. Once a plaintiff shows that a platform misuses this data to generate forbidden profits, it only makes sense that it will disgorge these profits to the original owners of the data. In this way, the platforms pay the full value of the data

---

260 Of course, the harms caused by platforms are borne by society at large and not exclusively by users. Therefore, it might be possible to allow nonusers to sue as individual plaintiffs under some circumstances. The Authors do not focus on this option here as this additional step seems largely unnecessary. The suggested claim is based on gains rather than harms, so the fact that nonusers are also harmed is of lesser importance. Additionally, limiting individual claims to users does not hinder litigation in any significant way as platform users are easy enough to come by.

261 *See* Gordon-Tapiero et al., *supra* note 2, at 647–51. It is estimated that in 2022, Facebook had more than 240 million users in the United States. Stacy Jo Dixon, *Number of Facebook Users in the United States From 2019 to 2028*, Statista (Jan. 30, 2024), https://www.statista.com/statistics/408971/number-of-us-facebook-users/ [https://perma.cc/FJK8-FHVZ].

262 For an explanation of such mechanisms, see, for example, Alon Harel & Alex Stein, *Auctioning for Loyalty: Selection and Monitoring of Class Counsel*, 22 Yale L. & Pol'y Rev. 69, 71 (2004).

263 *See id.*

264 *See id.* at 122 (quoting John C. Coffee, *Class Action Accountability: Reconciling Exit, Voice, and Loyalty in Representative Litigation*, 100 Colum. L. Rev. 370, 398 (2000)).

265 Glover, *supra* note 26, at 1154.

266 Shavell, *supra* note 259, at 359, 365.

they acquire, thus restoring the balance to the parties' bargain regarding an equal and fair exchange of assets.

In some cases, when distribution of class action awards to individual users is impracticable or inappropriate, courts can use cy pres relief as an alternative to traditional class action remedies.[267] Under this doctrine, courts can have the class defendant donate part of the award to a charitable cause related to the substance of the lawsuit.[268] In the context of platform personalization, such charitable causes might include digital literacy and online safety among vulnerable groups. Another venue for recovery would be for courts to order *fluid class recovery*.[269] Under this doctrinal alternative, courts can obligate the platform to award users with goods, services, future price reductions, or other monetary equivalents as a substitute to a monetary award.[270]

An additional solution for the implementation of these claims involves not using private plaintiffs at all but initiating claims through state actors who would be allowed to pursue an unjust enrichment claim against a platform in civil litigation. Such power can be used, for instance, by state attorneys general who have been known to utilize unjust enrichment claims in the name of public interest in other contexts.[271] Similarly, nongovernmental organizations dedicated to relevant issues such as media literacy could bring claims to courts. Such courses of action can prove useful when private plaintiffs are unwilling, or unable,[272] to bring their own claims; when informational advantages favor more public actors; or when the nature of the claim is such that a public plaintiff seems more appropriate to the court—for instance, when harms are spread over the population as a whole.

---

267 *See* Martin H. Redish, Peter Julian & Samantha Zyontz, *Cy Pres Relief and the Pathologies of the Modern Class Action: A Normative and Empirical Analysis*, 62 FLA. L. REV. 617, 634 (2010).

268 *Id.* ("In its current form as used in the federal courts, cy pres relief in class actions has involved the donation of a portion of the settlement or award fund to charitable uses which are in some loose manner connected to the substance of the case."). A 1972 student note pioneered the use of cy pres as a class action remedy. *See* Stewart R. Shepherd, Comment, *Damage Distribution in Class Actions: The Cy Pres Remedy*, 39 U. CHI. L. REV. 448, 448 (1972).

269 *See* Redish et al., *supra* note 267, at 662 (explaining the difference between cy pres relief and fluid class recovery); Gregory A. Hartman, Comment, *Due Process and Fluid Class Recovery*, 53 OR. L. REV. 225, 227 (1974).

270 *See* Redish et al., *supra* note 267, at 662.

271 In fact, attorneys general already have the power to do this. *See* Doug Rendleman, *Common Law Restitution in the Mississippi Tobacco Settlement: Did the Smoke Get in Their Eyes?*, 33 GA. L. REV. 847, 848 (1999) (describing such involvement of forty state attorneys general in the context of unjust enrichment tobacco litigation).

272 *See* Keith N. Hylton, *Litigation Costs and the Economic Theory of Tort Law*, 46 U. MIA. L. REV. 111, 113 (1991) (explaining how the costliness of litigation can bar plaintiffs from suing); Yotam Kaplan & Ittai Paldor, *Social Justice and the Structure of the Litigation System*, 101 N.C. L. REV. 469, 470–89 (2023) (highlighting the challenges private plaintiffs face in litigating against corporate litigants).

Thus, in some cases, private plaintiffs might enjoy an informational advantage or be more motivated to sue; in other cases, financial barriers may favor public plaintiffs. Overall, the flexibility in the identity of parties capable of initiating legal action against platforms will maximize deterrence and the probability that claims will arrive at court.

B. Predicted Outcomes

This proposal may seem to place a heavy burden on the activity of social media platforms. The Authors are not overly concerned, however, regarding the ability of platforms to survive despite these new burdens. As Paul Ohm aptly explains, "We couldn't kill [the internet] if we tried."[273] This Article anticipates that if the doctrine of unjust enrichment is applied to unjust platform profits as described above, platforms may choose to improve their way of doing business in one of several ways.

1. Updated Optimization Metrics

First, platforms may choose to adjust their optimization metrics. Engagement-based optimization metrics have had a detrimental effect on the type of content promoted to users.[274] There is no inherent reason to optimize for engagement other than maximizing potential profits from advertising. Once the incentive structure changes, and platforms can no longer expect to maintain profits generated by socially harmful practices, they are likely to find other optimization metrics. New optimization metrics will likely allow platforms to reach a new equilibrium whereby they are still able to generate profits but are more mindful of the way they generate them and the impact their activity has on society. Such an equilibrium will take time and experience to reach, but even the process of striving to achieve it is likely to have a positive societal impact.

Changing its algorithm's optimization metric is not new to Facebook. In the days following the 2020 U.S. presidential election, Facebook wanted to ensure it did not turn into an arena for people spreading false claims about the elections being stolen.[275] Facebook decided that its algorithm would prioritize news from sources deemed to be reliable by the platforms.[276] The news feed algorithm was therefore optimized

---

273 Paul Ohm, *We Couldn't Kill the Internet If We Tried*, 130 HARV. L. REV. F. 79, 85 (2016).

274 See *supra* Section I.C.

275 *See* Timberg et al., *supra* note 128.

276 Kevin Roose, *Facebook Reverses Postelection Algorithm Changes that Boosted News from Authoritative Sources*, N.Y. TIMES (Dec. 16, 2020), https://www.nytimes.com/2020/12/16/technology/facebook-reverses-postelection-algorithm-changes-that-boosted-news-from-authoritative-sources.html [https://perma.cc/7QJR-MW2N].

for an internal publisher score known as N.E.Q.—news ecosystem quality.[277] The change resulted in the prioritization of mainstream news outlets, such as *The New York Times* and *NPR*, and in substantially lower levels of promotion of disinformation.[278]

2. The Establishment of Civil Integrity Teams

This would not be the first time that leading social media platforms would take societal concerns into consideration in their content moderation decisions. Many view Facebook's involvement in the period leading up to the 2016 U.S. presidential election as problematic—allowing manipulative political ads, enabling the interference of foreign governments, and promoting disinformation to unsuspecting users.[279] Following this experience, Facebook arrived at the 2020 presidential election better prepared. It took several actions to ensure the integrity of the elections.[280] For example, Facebook's security team was entrusted with investigating and removing "coordinated networks of inauthentic accounts, Pages and Groups that [sought] to manipulate public debate."[281] It identified and removed fake accounts and took steps to secure "the accounts of elected officials, candidates and their staff."[282] The platforms also worked with "governments, law enforcement agencies, nonprofits, civil rights groups and other tech companies to stop emerging threats."[283] These efforts were coordinated by Facebook's civic integrity team.[284] The team's members were said to have subscribed to

---

277 *Id.*

278 *See id.*

279 *See* Alexis C. Madrigal, *What Facebook Did to American Democracy*, The Atlantic (Oct. 12, 2017), https://www.theatlantic.com/technology/archive/2017/10/what-facebook-did/542502/ [https://perma.cc/FND4-YVRQ]; Dipayan Ghosh & Ben Scott, *Facebook's New Controversy Shows How Easily Online Political Ads Can Manipulate You*, Time (Mar. 19, 2018, 12:38 PM), https://time.com/5197255/facebook-cambridge-analytica-donald-trump-ads-data/ [https://perma.cc/5QWD-6PA9]; Philip Bump, *All the Ways Trump's Campaign Was Aided by Facebook, Ranked by Importance*, Wash. Post (Mar. 22, 2018, 2:19 PM), https://www.washingtonpost.com/news/politics/wp/2018/03/22/all-the-ways-trumps-campaign-was-aided-by-facebook-ranked-by-importance/ [https://perma.cc/97NG-N27V; *see also* Mark Zuckerberg, Facebook (Sept. 27, 2017, 5:38 PM), https://www.facebook.com/zuck/posts/10104067130714241 [https://perma.cc/9968-3SNW] (denying President Trump's accusation that Facebook had always been against him).

280 Mike Isaac, *Facebook Moves to Limit Election Chaos in November*, N.Y. Times (Sept. 22, 2020), https://www.nytimes.com/2020/09/03/technology/facebook-election-chaos-november.html [https://perma.cc/ZNJ3-UPFB].

281 *Meta Policies and Safeguards for Elections Around the World*, Meta, https://about.facebook.com/actions/preparing-for-elections-on-facebook/ [https://perma.cc/RWZ7-2L9X].

282 *Id.*

283 *Our Approach to Elections*, Meta (Nov. 27, 2023), https://transparency.fb.com/features/approach-to-elections/ [https://perma.cc/6VZR-LX9G].

284 *See* Timberg et al., *supra* note 128.

an informal oath to "serve the people's interests first, not Facebook's."[285] Indeed, Facebook was very pleased with the way it had coped with the 2020 election, crowning its efforts to prevent manipulation of the election a success.[286] A month after the election, Facebook took actions to dissolve the civic integrity team, assigning its workers to other teams.[287] Frances Haugen, the Facebook whistleblower, was one of the Facebook workers who had high hopes for the civic integrity team especially after its success during the period leading up to the election.[288] She, along with other workers, was concerned that the dismantling of the team reflected an end to Facebook's willingness to forgo a certain level of profitability in favor of the protection of broader societal interests.[289] Five weeks later, Facebook users used the platform to both spread the conspiracy that the election had been rigged and stolen and to coordinate parts of the January 6th storming of the Capitol.[290] One way platforms may react if the doctrine of unjust enrichment is applied to harmful personalization may be to reinstate civic integrity teams where such existed or to establish similar bodies where they have not yet existed.

3. Tools to Combat Disinformation

There are several tools that platforms could use in order to curb the spread and prevalence of disinformation. One of the ways that Facebook combatted disinformation in the days following the 2020 presidential election was by closing groups promoting #stopthesteal and other elections-related conspiracy theories.[291] Platforms like Facebook offered personalized services such as suggesting groups to a user who may find their content interesting. An internal Facebook memo uncovered as part of *The Wall Street Journal's* Facebook Files shows that in August 2020 the platform was aware that 70 of the 100 top civic Facebook groups were full of "hate, bullying, harassment, [and] misinformation," and yet the platform continued recommending these groups to users.[292] Refraining from promoting hateful, dangerous groups

---

285 *See* Perrigo, *supra* note 129.

286 *See* Timberg et al., *supra* note 128.

287 *Id.*

288 Horwitz, *supra* note 61.

289 *See id.*; Timberg et al., *supra* note 128.

290 *See* Timberg et al., *supra* note 128; *see also* Sheera Frenkel, *The Storming of Capitol Hill Was Organized on Social Media*, N.Y. Times (Jan. 6, 2021), https://www.nytimes.com/2021/01/06/us/politics/protesters-storm-capitol-hill-building.html [https://perma.cc/H8UA-VPT5].

291 Shannon Bond, *Facebook Removes Pro-Trump Group Urging 'Boots on the Ground,'* NPR (Nov. 5, 2020, 2:57 PM), https://www.npr.org/2020/11/05/931794937/facebook-removes-pro-trump-group-urging-boots-on-the-ground [https://perma.cc/2C8U-VNUY].

292 Shannon Bond & Bobby Allyn, *How the 'Stop the Steal' Movement Outwitted Facebook Ahead of the Jan. 6 Insurrection*, NPR (Oct. 22, 2021, 9:50 PM), https://www.npr.org/2021/10/22/1048543513/facebook-groups-jan-6-insurrection [https://perma.cc/7DHZ-G6WG].

seems like a simple step that platforms could take in order to minimize their exposure to claims of unjust enrichment. Another way that these hateful groups can grow is by members sending out mass invites to their entire contact list.[293] Limiting the number of people each user can invite to such groups per day could limit their growth. Facebook implemented this recommendation in the period leading up to the 2020 election as it constrained the possible daily invitations to 100 and tightened this restriction as #stopthesteal gained traction after the election, limiting the permitted daily invitations to thirty.[294]

To ensure the integrity of the elections, Facebook utilized various "break glass" measures on the platform, most of which were removed following the end of the election process.[295] These measures were found to be effective in limiting the spread of disinformation. While reinstating them may indeed lower some users' engagement with the platforms, they are likely to be an effective way to overcome the societal costs that stem from the widespread dissemination of disinformation.[296] Disgorging profits generated by platforms based on the promotion of disinformation will change platform's incentives in a way that is likely to substantially lower the promotion of disinformation.

One of the meaningful ways that Facebook combats disinformation is by identifying, reviewing, and removing hate speech and other illegal content.[297] Facebook offers access to its platforms in 111 officially supported languages.[298] Other languages, not officially supported, are also in use on the platform.[299] At the same time, Facebook only has workers fluent in approximately fifty languages and its automated hate speech identification tools only operate in thirty languages.[300] Lack of fluency in some of the languages in which the platform operates could have dire consequences. For example, a *Reuters* report found that hate

---

293 *Id.*

294 *Id.*

295 Roose, *supra* note 276.

296 For a discussion of similar tools, see Erin Simpson & Adam Conner, *Fighting Coronavirus Misinformation and Disinformation*, Ctr. for Am. Progress (Aug. 18, 2020), https://www.americanprogress.org/article/fighting-coronavirus-misinformation-disinformation/ [https://perma.cc/VPK2-CYMZ].

297 *AI Advances to Better Detect Hate Speech*, Meta (May 12, 2020), https://ai.facebook.com/blog/ai-advances-to-better-detect-hate-speech/ [https://perma.cc/BW3W-2B7R]; *Detecting Violations*, Meta, https://transparency.fb.com/enforcement/detecting-violations/ [https://perma.cc/XT7G-KXQH].

298 Maggie Fick & Paresh Dave, *Facebook's Flood of Languages Leave it Struggling to Monitor Content*, Reuters (Apr. 23, 2019, 3:01 AM), https://web.archive.org/web/20230707164622/https://www.reuters.com/article/us-facebook-languages-insight-idUSKCN1RZ0DW [https://perma.cc/2QTJ-MAE9].

299 *Id.*

300 *Detecting Violations*, Meta, https://transparency.fb.com/enforcement/detecting-violations/ [https://perma.cc/XCE4-YMCE].

speech promoting ethnic cleansing posted on Facebook was unchecked because "the company's operation for monitoring content in Burmese was meagre."[301] In order to combat disinformation worldwide, platforms should ensure that they have workers who are fluent in all of the languages being used on the platform in different countries around the world. Focusing on removing disinformation appearing in English, while leaving users in other countries exposed to disinformation, sends a very problematic message in terms of Facebook's priorities. It is no doubt unjust to protect some users of the platforms while leaving the more vulnerable ones to defend themselves against the platform's incentive to promote engagement generating content. Reports show that Facebook workers gathered in 2019 to decide what election integrity measures would be implemented in each country.[302] Countries placed in tier zero would enjoy continuous monitoring of content by Facebook, while countries placed in tier three would only receive attention if certain content was reported to moderators.[303] Dal Yong Jin identifies this type of differential treatment as "platform imperialism," reflecting a reality whereby leading platforms have developed within a social and historical power context which they reinforce.[304] Thus, leading Western platforms act to ensure the safety and freedom of speech of users in certain countries, including the United States, while neglecting to provide the same level of protection for users in other countries.[305] Providing a different level of protection for the basic rights and safety of users based on their country of origin and the language they use creates a strong sense of injustice and discrimination, and any profits generated from doing so should be viewed as wrongful enrichment.

During the COVID-19 pandemic, disinformation regarding the virus, its dangers, and the vaccine and its potential effects were widespread. Numerous studies have pointed at YouTube as a source of both true and false information.[306] One study found that over twenty-seven percent of YouTube's most watched videos contained disinformation

---

301 Steve Stecklow, *Hatebook: Inside Facebook's Myanmar Operation*, Reuters (Aug. 15, 2018, 3:00 PM), https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/ [https://perma.cc/7CSK-JUYA].

302 Casey Newton, *The Tier List: How Facebook Decides Which Countries Need Protection*, Verge (Oct. 25, 2021, 7:00 AM), https://www.theverge.com/22743753/facebook-tier-list-countries-leaked-documents-content-moderation [https://perma.cc/X4LJ-T2QG].

303 *Id.*

304 Dal Yong Jin, *The Construction of Platform Imperialism in the Globalization Era*, 11 Commc'n, Capitalism & Critique 145, 146 (2013).

305 *See* Sara Bannerman, *Platform Imperialism, Communications Law and Relational Sovereignty*, New Media & Soc'y, 2022, at 3.

306 *See, e.g.*, Charles E. Basch, Corey H. Basch, Grace C. Hillyer, Zoe C. Meleo-Erwin & Emily A. Zagnit, *YouTube Videos and Informed Decision-Making About COVID-19 Vaccination: Successive Sampling Study*, JMIR Pub. Health Surveillance, May 2021, at 1.

regarding the pandemic.[307] The platform's policy regarding COVID-19 related disinformation included two main tools. The first was the removal of "content that falsely alleges that approved vaccines are dangerous and cause chronic health effects."[308] The second step that the platform took was to accompany COVID-19 and vaccine-related content with links to where users could access reliable information, namely the Center for Disease Control's and the World Health Organization's websites.[309] Reinstating such tools may help platforms in an unjust enrichment claim brought against them.

The law should not permit platforms to continue profiting from the promotion of disinformation that has a real potential to harm the health and safety of society and individuals within it.

Conclusion

The platform crisis has pushed democracies toward the edge of the precipice. As long as harmful personalization practices generate profits for platforms, they will continue implementing them. Something must be done soon if we are to pull ourselves back and survive this crisis. A fundamental change to platforms' incentive structure is required. This Article proposes this change through the law of unjust enrichment by removing platforms' gains when they are obtained in ways that are clearly socially harmful. This solution is not only necessary as a matter of policy, but also follows naturally from existing doctrines of the law of unjust enrichment.

This proposal is a game changer in terms of the ability of the legal system to contend with the current crisis. The proposed framework enjoys several significant advantages. First, the law of unjust enrichment can assure effective deterrence by removing the profits platforms obtain through their wrongful activities. Second, the harms of platforms' practices are often difficult to identify and measure. Therefore, harm-based remedies are often unavailable or impossible to operate. Conversely, platform profits are all too real and much easier to measure. Third, the doctrinal tests embodied in the law of unjust enrichment offers the level of flexibility required to regulate the ever-changing landscape of platform activity. Fourth, the law of unjust enrichment draws on the

---

307 Heidi Oi-Yee Li, Adrian Bailey, David Huynh & James Chan, *YouTube as a Source of Information on COVID-19: A Pandemic of Misinformation?*, BMJ Glob. Health, May 2020, at 1, 3.

308 The YouTube Team, *Managing Harmful Vaccine Content on YouTube*, YouTube: Off. Blog (Sept. 29, 2021), https://blog.youtube/news-and-events/managing-harmful-vaccine-content-youtube/ [https://perma.cc/KCL8-XVPE].

309 *Topical Context in Information Panel*, YouTube Help, https://support.google.com/youtube/answer/9004474?hl=en [https://perma.cc/KUL9-YXL2].

comparative advantages of diverse actors, including private plaintiffs, courts, regulators, and experts, and can therefore generate effective and informed legal action. The proposal detailed in this Article has the power to meaningfully change the financial incentives of platforms, thereby protecting individuals and society as a whole.