THE CASE FOR ESTABLISHING A COLLECTIVE PERSPECTIVE TO ADDRESS THE HARMS OF PLATFORM PERSONALIZATION

Ayelet Gordon-Tapiero,[*] *Alexandra Wood,*[**] *& Katrina Ligett*[***]

ABSTRACT

Personalization on digital platforms drives a broad range of harms, including misinformation, manipulation, social polarization, subversion of autonomy, and discrimination. In recent years, policymakers, civil society advocates, and researchers have proposed a wide range of interventions to address these challenges. In this Article, we argue that the emerging toolkit reflects an individualistic view of both personal data and data-driven harms

[*] Research Fellow, the Federmann Cyber Security Center, Postdoctoral Research Fellow, the Rachel and Selim Benin School of Computer Science and Engineering, the Hebrew University of Jerusalem.

[**] Fellow, Berkman Klein Center for Internet & Society at Harvard University.

[***] Associate Professor, the Rachel and Selim Benin School of Computer Science and Engineering, the Hebrew University of Jerusalem.

The authors would like to thank Gordon Berlin, Michael Birnhack, Ella Corren, Dafna Dror, Elizabeth Edenberg, Alon Harel, Daniel Ho, Neta Livneh, Kobbi Nissim, Gideon Parchomovsky, Matt Prewitt, Amnon Reichman, Gal Shahaf, Thomas Streinz, Bo Waggoner, and Eyal Zamir for their invaluable insights. In addition, we are grateful for useful comments provided by the participants in a working group on data co-ops hosted by the Simons Institute for the Theory of Computing in Spring 2019; the Weizmann University program on Societal Concerns in Algorithms and Data Analysis; the Tel Aviv University Law Workshop in Information Technology and Law; a seminar held at the Berkman Klein Center for Internet & Society at Harvard University; the Harvard Privacy Tools Project; the Boston University Cyber Security, Law, and Society Alliance; the Colloquium of Microsoft Research New England; the JerusML Show; a series of workshops on data co-ops organized at Georgetown University and the Hebrew University of Jerusalem; Intel Jerusalem Techweek; the Tel Aviv University Workshop on AI, Law and Agency in the Age of Machine Learning; the RadicalXChange Annual Conference; the colloquium of the Max Planck Institute for Software Systems; the DIMACS Workshop on the Co-Development of Computer Science and Law; the Privacy in Machine Learning (priML)/Privacy-Preserving Machine Learning (PPML) Workshop; the 8th Privacy, Cyber and Technology Workshop at Tel Aviv University; and the Cyber Forum at the Haifa Center for Cyber, Law and Policy. This work has been supported by a workshop grant from the Simons Foundation, a gift to the McCourt School of Public Policy and Georgetown University, Simons Foundation Collaboration 733792, and Israel Science Foundation (ISF) grants 1044/16 and 2861/20. Part of Ligett's work was done while the author was visiting the Simons Institute for the Theory of Computing.

Electronic copy available at: https://ssrn.com/abstract=4105443

that will likely be inadequate to address growing harms in the global data ecosystem. We maintain that interventions must be grounded in an understanding of the fundamentally collective nature of data, wherein platforms leverage complex patterns of behaviors and characteristics observed across a large population to draw inferences and make predictions about individuals.

Using the lens of the collective nature of data, we evaluate various approaches to addressing personalization-driven harms under current consideration. This lens also allows us to frame concrete guidance for future legislation in this space and advocate meaningful transparency that goes far beyond current transparency proposals. We offer a roadmap for what meaningful transparency must constitute: a collective perspective providing a third party with ongoing insight into the information gathered and observed about individuals and how it correlates with any personalized content they receive, across a large, representative population. These insights would enable the third party to understand, identify, quantify, and address cases of personalization-driven harms. We discuss how such transparency can be achieved without sacrificing privacy and provide guidelines for legislation to support the development of such transparency.

CONTENTS

Electronic copy available at: https://ssrn.com/abstract=4105443

INTRODUCTION

Platforms' ability to personalize content for each of their users has recently given rise to several controversial cases including the Facebook-Cambridge Analytica data scandal;[1] the emotional contagion experiment on influencing Facebook users' moods;[2] research finding that leading platforms discriminate in their presentation of job and housing ads on the basis of race, gender, and age;[3] and, most recently, the Wall Street Journal's investigative

[1] *See* Nicholas Confessore, *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far,* N.Y. TIMES, Apr. 4, 2018, https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

[2] *See* Adam D.I. Kramer, Jamie E. Guillory & Jeffrey T. Hancock, *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks*, 111 PRO. NAT'L ACAD. SCI. 8788 (2014).

[3] *See, e.g.,* Basileal Imana, Aleksandra Korolova & John Heidemann, *Auditing for Discrimination in Algorithms Delivering Job Ads*, PROC. WEB CONF. 3767, 3769 (2021) (demonstrating that presentation of ads on Facebook, Google, and LinkedIn can be skewed by gender); Alexia Fernández Campbell, *Job Ads on Facebook Discriminated Against Women and Older Workers, EEOC Says*, VOX, Sep. 25, 2019, https://www.vox.com/identities/2019/9/25/20883446/facebook-job-ads-discrimination (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* (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3570076 (finding that GoogleAds presented users who had previously searched for Black names with ads for disadvantageous jobs compared to users who had previously searched for White names).

Electronic copy available at: https://ssrn.com/abstract=4105443

reporting on The Facebook Files.[4] Following the testimony of Facebook whistleblower Frances Haugen before Congress in October 2021, a rare bipartisan response signified that "Facebook and Big Tech are facing a Big Tobacco moment,"[5] with voices in the United States and around the world calling for stronger regulation of platforms.[6]

For an introduction to platforms' approaches to personalization, see Kimberly Rhum, *Information Fiduciaries and Political Microtargeting: A Legal Framework for Regulating Political Advertising on Digital Platforms*, 115 N. W. L. REV. 1829, 1831 (2021) (detailing how a variety of platforms offer their users personalized experiences).

[4] Jeff Horowitz, *The Facebook Files*, WALL ST. J., Oct. 1, 2021, https://www.wsj.com/articles/the-facebook-files-11631713039.

[5] Cecilia Kang, *Lawmakers See Part to Rein in Tech, But It Isn't Smooth*, N.Y. TIMES, Oct. 9, 2021, https://www.nytimes.com/2021/10/09/technology/facebook-big-tobacco-regulation.html.

[6] *See* Adam Satariano, *Facebook Hearing Strengthens Calls for Regulation in Europe*, N.Y. TIMES, Oct. 6, 2021, https://www.nytimes.com/2021/10/06/technology/facebook-european-union-regulation.html.

Electronic copy available at: https://ssrn.com/abstract=4105443

Economic, social, and cultural activities are increasingly mediated by platforms,[7] representing a shift "from industrial to information capitalism."[8] As the process of digitization has enabled increased datafication, platforms' power and control over the modern marketplace for economic, social, and cultural interactions have grown.[9] To manage and leverage the growing amount of electronic data they possess, platforms have developed and implemented artificial intelligence and machine learning algorithms, which,

[7] Several definitions of the term *platform* have been offered in the literature. For example, Lina Kahn emphasizes platforms' role as intermediaries of economic activities, likening them to bank holding companies. Lina M. Kahn, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710, 795 (2017). Other definitions focus on the fact that platforms do not only mediate economic transactions, but "in a broader social sense of comprising the basic infrastructure of modern society." K. Sabeel Rahman, *The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept*, 39 CARDOZO L. REV. 1621, 1641 (2018). Perhaps one of the most important areas in which platforms have had a transformative role is that of data production and collection. Indeed, Cohen recognizes that platforms' greatest interest lies in "data extracted from people as they invest, work, operate businesses, socialize, and engage in innumerable other activities." JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 38 (2019). *See also* Priscilla M. Regan, *A Design for Public Trustee and Privacy Protection Regulation*, 44 SETON HALL LEGIS. J. 487, 496 (2020) ("It is widely recognized that the business models of large internet companies rely upon the collection, use, and analysis of personal information").

In this Article, we build on Cohen's recognition of the central role of data in the business models and activities of platforms, using the term to refer to entities that collect, store, process, analyze, or act upon data pertaining to individuals (for example, in the provision of content, services, recommendations, or ads), and whose presence is primarily in the digital realm. We use the term *users* to denote individuals who use the services of the platforms. The term *individuals* describes people (who have not necessarily signed up to use a certain platform or agreed to its terms of service). Finally, the term *data ecosystem* refers to platforms, individuals, and any other entities participating in exchanging, transacting, and acting on data pertaining to individuals. *See* Salome Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573, 612 (2021) (noting "the combination of relational and aggregate effects from data production drive companies to collect as much data as possible from data subjects").

[8] COHEN, *id.* at 7.

[9] *See* SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 186-7 (2020); COHEN, *supra* note 7, at 28 (detailing an example of how platforms became involved in the field of consumer finance).

Electronic copy available at: https://ssrn.com/abstract=4105443

in turn, demand large volumes of data as inputs.[10] Common across platforms' various business models is a strong incentive to collect and analyze massive quantities of data about individuals — and to use this information to present individuals with personalized content.[11]

Platforms harness their ability to capture, analyze, and act upon data on the behavior of large groups; detect patterns of behavior and previously unanticipated clusters of users; make predictions about how individuals and groups of individuals will respond to personalized content; infer deeply personal attributes that an individual has not expressly disclosed; and act upon these predictions and inferences.[12] Such personalization, i.e., the ability to show each user content specifically chosen for them, can benefit users, but it also contributes to a broad range of data-based harms, including

[10] *See* Josep Lluis Berral-Garcia, *A Quick View on Current Techniques and Machine Learning Algorithms for Big Data Analytics*, INT'L CONF. TRANSPARENT OPTICAL NETWORKS, IEEE (2016) (explaining that in order to manage big data the development of machine learning algorithms is necessary); Jafar Alzubi, Anand Nayyar & Akshi Kumar, *Machine Learning from Theory to Algorithms: An Overview*, J. PHYSICS, CONF. SERIES. 1, 13 (2018) (observing that "machine learning algorithms require large volumes of data to be accurate and efficient").

[11] *See* Julie E. Cohen, *Law for the Platform Economy*, 51 U. C. DAVIS L. REV. 133 (2017) (discussing how ongoing collection of large amounts of data is an important part of platforms' market power); Shoshana Zuboff, *Big Brother: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 J. INFO. TECH. 75 (2015) (explaining that the ability to collect large amounts of data is a significant part of the *surveillance economy*);
Hal R. Varian, *Computer Mediated Transactions*, 100 AMERICAN ECON. REV.: PAPERS & PROC. 1 (2010) (identifying relatively early on in the development of the internet that facilitating personalization was one of the substantial impacts of computer mediated transactions); Brent Mittelstadt, *Auditing for Transparency in Content Personalization Systems*, 10 INT'L J. COMM 4991 (2016) ("Content personalization systems display information tailored to individual users, often based on perceived preferences or past behaviors.").

[12] *See* Eduard Fosch-Villaronga, Adam Poulsen, Roger Andre Søraa & Bart Custers, *A Little Bird Told Me Your Gender: Gender Inferences in Social Media*, 58 INFO. PROCESS & MGMT. 1 (2021) (demonstrating that platforms can infer an individual's gender even when they have not provided it).

Electronic copy available at: https://ssrn.com/abstract=4105443

misinformation,[13] manipulation,[14] social polarization,[15] subversion of autonomy,[16] and discrimination.[17]

As a consequence, early optimism that the Internet would evolve to be a "liberating and democratic social force"[18] has all but faded away, and in recent years, policymakers, civil society advocates, and researchers around the world have increasingly turned their attention to the challenges facing the data ecosystem.[19]

---

[13] *See* Ashley Smith-Roberts, *Facebook, Fake News, and the First Amendment*, 95 DENVER L. R. F. 118 (2018).

[14] *See* Zeynep Tufekci, *Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 COLO. TECH. L.J. 203 (2015).

[15] *See* Christopher A. Bail et al., *Exposure to Opposing Views on Social Media can Increase Political Polarization*, 37 PROC. NAT'L ACAD. SCI. 9216 (2018).

[16] *See* Daniel Susser, Beate Roessler & Helen Nissenbaum, *Technology, Autonomy and Manipulation*, 8 INTERNET POL'Y REV. 1 (2019).

[17] For example, U.S. anti-discrimination laws prohibit discrimination in housing and employment advertising. *See* 42 U.S.C. § 804; 42 U.S.C. § 2000e. Section 804 of the Fair Housing Act served as the basis for the U.S. Department of Housing and Urban Development's charge of discrimination against Facebook in 2019, alleging discrimination in the presentation of ads for housing on the platform U.S. DEP'T HOUS. & URB. DEV., Charge of Discrimination, FHEO No. 01-18-0323-8 (2019). Section 2000e of Title VII of the Civil Rights Act of 1964 served as the basis for a decision by the U.S. Equal Employment Opportunity Commission finding that seven employers had violated federal law when advertising jobs on Facebook in a way that excluded women and/or older workers from seeing the ads. ACLU, *In Historic Decision on Digital Bias, EEOC Finds Employers Violated Federal Law when they Excluded Women and Older Workers from Facebook Ads*, Sep. 25, 2019, https://www.aclu.org/press-releases/historic-decision-digital-bias-eeoc-finds-employers-violated-federal-law-when-they (reporting on the decision); U.S. Equal Emp. Opportunity Comm'n, Letters of Determination, July 5, 2019, available at https://www.onlineagediscrimination.com/sites/default/files/documents/eeoc-determinations.pdf [https://perma.cc/4LNE-F3N5]. Researchers have also demonstrated that numerous platforms present housing and employment ads in a discriminatory manner. *See, e.g.,* Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove & Aaron Rieke, *Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes*, 3 PROC. ACM ON HUM.-COMPUT. INTERACTION 1, 1 (2019) (observing significant skews in the presentation of ads for housing and employment along gender and racial lines); Imana et al., *supra* note 3 at 3769 (demonstrating that presentation of ads on Facebook, Google, and LinkedIn can be skewed by gender).

[18] ZUBOFF, *supra* note 9, at 67. *See* Amy Kapczynski, *The Law of Informational Capitalism*, 129 YALE L.J. 1460, 1460 (2020).

[19] Explosive growth in the global data ecosystem has led to the recent adoption of a number of data protection and consumer privacy laws. *See, e.g.,* Regulation 2016/679 of the EUR. PARL. & COUNCIL of Apr. 27, 2016 on the protection of natural persons with regard to the

Electronic copy available at: https://ssrn.com/abstract=4105443

In this Article, we argue that the emerging toolkit reflects an individualistic view of personal data and data-driven harms and that such a framing will likely fail to adequately address the harms stemming from platform personalization. We assert that, instead, interventions must be grounded in an understanding of the fundamentally collective nature of data,[20] i.e., the personalized content that one person receives is strongly

processing of personal data and on the free movement of such data [hereinafter GDPR]; California Privacy Rights Act of 2020, CAL. CIV. CODE § 1798.100 [hereinafter CPRA]; California Consumer Privacy Act, CAL. CIV. CODE § 1798.100 [hereinafter CCPA]; Virginia Consumer Data Protection Act, VA. CODE ANN. § 59.1; Colorado Privacy Act, COLO. REV. STAT. § 6-1.

In the United States, several legislative proposals have targeted the harms stemming from platform personalization. *See, e.g.*, the Honest Ads Act, H.R. 4077, 115th Cong. (1st Sess. 2017) [hereinafter Honest Ads Act], the Deceptive Experiences to Online Users Reduction (DETOUR) Act, S. 1084, 116th Cong. (2019-2020) [hereinafter DETOUR Act] ("To prohibit the usage of exploitative and deceptive practices by large online operators and to promote consumer welfare in the use of behavioral research by such providers"); the Social Media Addiction Reduction Technology Act (SMART) Act, S. 2314, 116th Cong. (2019-2020) [hereinafter SMART Act]; the Filter Bubble Transparency Act, 116th Cong. (2019-2020) [hereinafter FBTA]; the Children and Media Research Advancement Act (CAMRA) Act, S. 971, 117th Cong. (2021-2022); the Protecting Americans from Dangerous Algorithms Act, H.R. 2154, 117th Cong. (2021-2022) [hereinafter Protecting Americans from Dangerous Algorithms Act]; the Justice Against Malicious Algorithms Act of 2021, H.R. 5596, 117th Cong. (2021-2022); the Health Misinformation Act of 2021, S. 2448, 117th Cong. (2021-2022); the Social Media Disclosure and Transparency (DATA) Act, H.R. 3451, 117th Cong. (1st Sess. 2021) [hereinafter Social Media DATA Act]; the Platform Accountability and Transparency Act (PATA), S. ____ , 117th Cong. (1st Sess. 2021) [hereinafter PATA]; the Algorithmic Justice and Online Platform Transparency Act, S. 1896, 117th Cong. (1st Sess. 2021).

In Europe, several initiatives to address the challenges of personalization have been introduced. *See, e.g*., The Eur. Comm'n, Proposal for a Regulation of the Parliament and of the Council on a Single Market for Digital Services and amending Directive 2000/31/EC', COM/2020/825 [hereinafter Digital Services Act] (aiming to "establish a powerful transparency and a clear accountability framework for online platforms"); Eur. Comm'n, EU Code of Practice on Discrimination (2018), https://ec.europa.eu/newsroom/dae/document.cfm?docid=54454 (adopting self-regulatory standards to combat disinformation).

[20] *See* Martin Tisne, *The Data Delusion: Protecting Individual Data Isn't Enough when the Harm Is Collective*, LUMINATE 2 (2020) (The collective nature of big data means people are more impacted by other people's data than by data about them. Like climate change, the threat is societal *and* personal"); Regan, *supra* note 7, at 501 ("There is no question that regulators are struggling and not doing very well in this struggle.").

Electronic copy available at: https://ssrn.com/abstract=4105443

driven by rich data gathered about others around the globe.[21] Many platform-driven challenges such as social polarization and discrimination cannot be defined with respect to one isolated individual; these harms and their definitions, as well as the ability to detect them, inherently exist within a broader social context.[22] Furthermore, the only parties that may currently possess a picture of this personalization landscape are the platforms themselves.[23] Carefully constructed experiments have demonstrated that platforms induce discriminatory personalization of certain content, such as ads for employment.[24] Such experiments, however, are inherently limited in scope, and are able to identify only instances of the particular harm they were looking for at the time they were conducted. They do not provide the deeper transparency that society needs. Adequate transparency furthermore requires far more than disclosing ad targeting criteria or ad funding details as in the Honest Ads Act,[25] creating databases of ads divorced from the personal

[21] *See* Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, COLUM. BUS. L. REV. 494, 502 (2019).
[22] *See* Simon A. Levin, Helen V. Milner & Charles Perrings, *The Dynamics of Political Polarization*, 118 PROC. NAT'L ACAD. SCI. 1 (2021) (acknowledging that phenomena such as polarization "are inherently systems-level phenomena, involving interactions among multiple component parts and the emergence of broader scale features").
[23] *See* Yochai Benkler, *Degrees of Freedom, Dimensions of Power*, 145 DAEDALUS 18, 23 (2016) ("Big data collection and processing, combined with ubiquitous sensing and connectivity, create extremely powerful insights on mass populations available to relatively few entities.").
[24] *See, e.g.,* Muhammad Ali, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove & Aaron Rieke, *Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging* 1, https://arxiv.org/pdf/1912.04255.pdf (2019) (finding that "Facebook preferentially exposes users to political advertising that it believes is relevant for them"); Imana et al., *supra* note 3, at 3767 (demonstrating that presentation of ads on Facebook, Google, and LinkedIn can be skewed by gender); Amit Datta, Michael Carl Tschantz & Anupam Datta, *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination*, PROC. PRIV. ENHANCING TECH. 92, 92 (2015) (demonstrating that changing one's self-reported gender influences the job ads one sees).
[25] The Honest Ads Act, *supra* note 19.

Electronic copy available at: https://ssrn.com/abstract=4105443

information of those who received them as in the Digital Services Act,[26] or focusing primarily on ads as in the Social Media DATA Act.[27]

Without meaningful, effective transparency, society lacks the essential tools to properly understand the role that personalization plays in generating and amplifying various harms. At present, there is uncertainty regarding even the most basic questions, such as whether personalization is contributing to polarization or defusing it.[28] We offer a roadmap for what meaningful transparency must constitute: providing a third party with ongoing insight into the information gathered and observed about individuals and how it correlates with any personalized content they receive, across a large, representative population. We discuss how such transparency can be achieved without sacrificing privacy and provide guidelines for regulation to support the development of such transparency.

The Article proceeds as follows. Part I describes the structure of the data ecosystem, explains the financial incentives driving platforms' extensive data collection, and introduces novel terminology that captures the different flows of content between users and platforms. It also highlights the various ways in which data is collective and demonstrates how information about one person can allow a platform to learn about another person. Part II uses this lens of the collective nature of data to help analyze various regulatory and technical approaches designed to address personalization-driven harms. In Part III, we present design principles that can facilitate effective intervention. We advocate for meaningful transparency, generating a collective perspective that would allow a third party to view the data of large groups of

[26] Digital Services Act, *supra* note 19.
[27] Social Media DATA Act, *supra* note 19.
[28] *See* Levi Boxell, Matthew Gentzkow & Jesse M. Shapiro, *Is the Internet Causing Political Polarization? Evidence from Demographics*, NBER WORKING PAPER SERIES (2017) (demonstrating that the age group exhibiting the highest level of polarization was the group aged 75 and older, i.e., the age bracket with the least exposure to the internet and social media); *see also* Bail et al., *supra* note 15.

Electronic copy available at: https://ssrn.com/abstract=4105443

users and offer ways regulation could facilitate the creation of such a perspective.

I. THE STRUCTURE OF THE DATA ECOSYSTEM

In this Part, we provide an overview of the structure of the data ecosystem and the incentives driving platforms' activities.[29] In particular, their business models have created powerful incentives – and capabilities – for platforms to design their services, content, and interfaces to increase opportunities for impactful personalized advertising, thereby boosting profitable revenue streams.[30] We introduce terminology describing the flows of content between users and platforms and how these flows create a feedback loop: data collected by platforms serves as a basis for personalizing content for users, whose activity then generates more data to be collected by platforms. In addition, we demonstrate why it is critical to recognize the collective nature of data when considering the suitability of interventions to address personalization-driven harms.

A. The Data Ecosystem's Outgoing and Incoming Vectors

[29] Platforms' business models vary based on numerous criteria, such as whether individuals pay to access the service, to what extent advertising is a significant part of the platform's revenue, what type of data the platform gathers, which parties it shares data with, what information services the platform provides, and how personalized the offered services are. In this Article, we refer to all platforms as defined in *supra* note 7, regardless of their business model.

[30] One of the byproducts of platforms' ability to personalize ads and other content – indeed of informational capitalism as a broad phenomenon – is a deepening of social inequality. Platforms have amassed power while society has seen the emergence of a "seemingly permanent economic underclass." COHEN, *supra* note 7, at 180; *see also* Tim Berners-Lee, *One Small Step for the Web,* MEDIUM, Sept. 29, 2018, https://medium.com/@timberners_lee/one-small-step-for-the-web-87f92217d085 (observing that "for all the good we've achieved, the web has evolved into an engine of inequity and division; swayed by powerful forces who use it for their own agendas").

Electronic copy available at: https://ssrn.com/abstract=4105443

In the data ecosystem, information flows between users and platforms in two directions. First, data flows from users to platforms in what we call the *outgoing vector*. Along the outgoing vector, platforms collect vast quantities of data about users and their activities,[31] including interactions each user has directly with the platform (e.g., groups they belong to, and pages and other content they "like"), interactions between users (e.g., commenting on a friend's post, retweeting, and sharing a video), and users' online activity outside the platform (e.g., which other web sites a user has visited).[32] In some cases, platforms also collect information about users' offline activity that is

[31] *See* Datta et al., *supra* note 24, at 92 ("Colossal amounts of collected data are used, sold, and resold for serving targeted content, notably advertisements, on websites."); 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); Shira Ovide, *What's Behind the Apple-Facebook Feud?,* N.Y. TIMES, Apr. 26 2021, https://www.nytimes.com/2021/04/26/technology/apple-facebook-feud.html ("Currently, Facebook and companies like it track the ways people use their phones, picking up bits of information such as how often they open their yoga app and what they buy at Target. Facebook then uses that information to help companies target their ads."); 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*, PROC. 1ST CONF.ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 1, 3 (2018) ("Facebook gathers and infers several hundreds of attributes for all of its users."); 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 your friends with.").

[32] When a user signs into a third-party service with their Facebook account, Facebook is made aware of their activity, even though it takes place outside the Facebook platform. Additionally, when a Facebook user visits a site with the 'like' button embedded in it, Facebook collects information about that visit regardless of whether the user clicked the 'like' button. *See* Jonathan R. Mayer & John C. Mitchell, *Third-Party Web Tracking: Policy and Technology*, 2012 IEEE SYMP. ON SEC. & PRIV. 413, 419 (2012); Dina Srinivasan, *The Antitrust Case Against Facebook: A Monopolist's Journey Towards Pervasive Surveillance in Spite of Consumer's Preference for Privacy*, 16 BERKELEY BUS. L. J. 39, 41 (2019). Additionally, Google keeps track of news articles that its users read, even if they are not accessed via a Google search. *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 ("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.").

Electronic copy available at: https://ssrn.com/abstract=4105443

provided by their devices (such as their location data)[33] or by third parties, including information about users' shopping habits, credit scores, public records such as census data, voter registration data, and more.[34] Platforms collect and analyze this data in order to draw a detailed profile about each user, and, at times, they make it available to third parties as well.[35] In this

---

[33] *See* Chen, *id.* ("On some days, [Facebook] even logged my locations, like when I was at a hospital two years ago or when I visited Tokyo last year."); Irfan Faizullabhoy & Aleksandra Korolova, *Facebook's Advertising Platform: New Attack Vectors and the Need for Interventions*, Computing Research Repository (2018) (unpublished manuscript), https://arxiv.org/abs/1803.10099 ("Social media websites such as Facebook, Google, and Pinterest record and learn from user behavior, [...] such as location."); John Herrman, *Google Knows Where You've Been, But Does It Know Who You Are?*, N.Y. TIMES, Sep. 12, 2018, https://www.nytimes.com/2018/09/12/magazine/google-maps-location-data-privacy.html ("Some Google apps automatically store time-stamped location data without asking.").

[34] *See* Ali et al., *supra* note 24, at 4 ("Facebook receives information from a variety of sources beyond the Facebook website and app, including Facebook Pixel tracking, app data sharing, third-party data brokers, and location data.") (citations omitted); 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*, THE WORLD WIDE WEB CONFERENCE 1920, 1920 (2019) ("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."); 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."); Kalev Leetaru, *The Data Brokers So Powerful Even Facebook Bought Their Data - But They Got Me Wildly Wrong,* FORBES, Apr. 5, 2018, https://www.forbes.com/sites/kalevleetaru/2018/04/05/the-data-brokers-so-powerful-even-facebook-bought-their-data-but-they-got-me-wildly-wrong ("In essence, Facebook recognized that many of the most useful data points on our daily lives come not from the utopian image of perfection we project on Facebook, but from the actual mundane reality of our daily lives, from what we purchase at the grocery store to where we live to our financial status."); Kashmir Hill, *Facebook Is Tracking What Users Buy In Stores To See Whether Its Ads Work,* FORBES, Sep. 26, 2012, https://www.forbes.com/sites/kashmirhill/2012/09/26/facebook-is-tracking-what-users-buy-in-stores-to-see-whether-its-ads-work.

[35] *See* Gabriel J. X. Dance, Nicholas Confessore & Michael LaForgia, *Facebook Gave Device Makers Deep Access to Data on Users and Friends,* N.Y. TIMES, June 3, 2018, https://www.nytimes.com/interactive/2018/06/03/technology/facebook-device-partners-users-friends-data.html ("Facebook has reached data-sharing partnerships with at least 60 device makers — including Apple, Amazon, BlackBerry, Microsoft and Samsung — over the last decade, starting before Facebook apps were widely available on smartphones, company officials said."); *see also* Venkatadri et al., *supra* note 34, at 1920; Nizan Geslevich Packin, *Show Me the (Data About the) Money!,* 5 UTAH L. REV. 1277, 1310 (2020) ("FinTech apps collect more data than needed, save it in an unsafe way, and sell it to third-parties.").

Electronic copy available at: https://ssrn.com/abstract=4105443

Article, we view privacy as predominantly an outgoing-vector concern, related to mitigating the platform-mediated flow of data pertaining to individuals.[36]

In the other direction, information flows from platforms to users, in what we call the *incoming vector*. This encompasses all content that platforms present or suggest to users based on the detailed profile that the platform has created about them,[37] including sending notifications, resurfacing an old post as a memory, compiling photos and other user-generated content into custom videos, suggesting groups to join, and presenting posts or videos to users in their feed (including the order in which they are presented), as well as providing suggestions for other content they may be interested in (e.g., news articles, physical gatherings, and other users to connect with). Personalization along the incoming vector is designed to increase user engagement and time spent on the platform. The more time a user spends interacting with the platform, the more data the platform collects, allowing it to increasingly present more accurately tailored personalized content.

Whereas platforms derive enormous profits from users' data, users do not enjoy a portion of these financial benefits. Scholars, activists and technologists have proposed changes in data governance to overcome this imbalance of power online; two central suggestions include treating data as property and providing "fundamental-rights protections to data as an extension of personal selfhood." *See* Viljoen*, supra* note 7, at 617.

[36] While we observe that approaches to privacy in practice tend to focus primarily on addressing outgoing-vector concerns, we recognize that some dimensions of privacy and data protection, such as the principles of purpose limitation and data minimization, among others, are also relevant to incoming-vector concerns. *See* discussion *infra* Section II.A.

[37] This information flow also includes the order in which the newsfeed or timeline is presented, and content such as compiling photos and other content into a friendship anniversary movie, suggestions to join groups, and more. *See* Sandra Wachter, *Affinity Profiling and Discrimination by Association in Online Behavioral Advertising*, 35 BERKELEY TECH. L.J. 367, 369 (2020) (describing platforms' ability to personalize content for users based on platforms' knowledge of users' personal attributes).

Electronic copy available at: https://ssrn.com/abstract=4105443

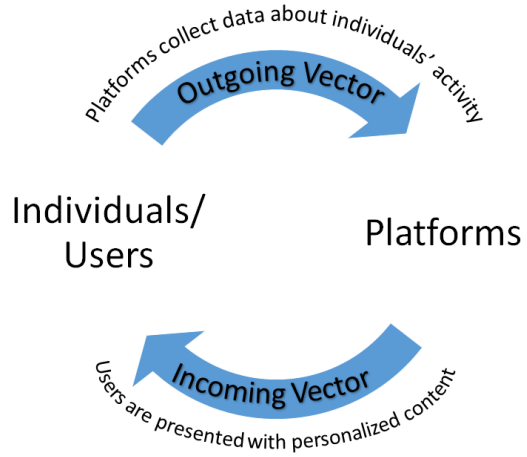

Figure 1. Schematic of the flows of information via the outgoing and incoming vectors.

As illustrated in Figure 1, the outgoing and incoming vectors create a feedback loop: data collected along the outgoing vector is analyzed as a basis for content personalization along the incoming vector, and individuals' interactions with personalized content presented to them along the incoming vector then generate more information along the outgoing vector for the platforms to collect and analyze.

B. The Collective Nature of Data

In this Section, we describe the fundamentally collective nature of data within the data ecosystem, whereby data about one individual can enable platforms to learn about another individual, and patterns of data detected across groups of users also provide insight into the behavior or characteristics of others. We argue that recognition of the collective nature of data should

Electronic copy available at: https://ssrn.com/abstract=4105443

inform any intervention to address the harms stemming from incoming-vector personalization.[38]

Each platform user is associated with an extensive record of behavior, such as searches conducted, links clicked, posts liked, messages sent, photos posted, social connections formed, and more.[39] Although there is a tendency to think of such data as belonging to a single user,[40] the reality is much blurrier,[41] as a message sent pertains to both the sender and to the recipient[42] and a photo in which a friend is tagged pertains both to the poster and to the subject.[43] Similarly, a Google search for information about a rare genetic

[38] *See* discussion *infra* in this Section on the different ways in which the information about one user can teach a platform about another user. *See also* Lars Backstrom, Cynthia Dwork & Jon Kleinberg, *Wherefore Art Thou R3579X?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography*, PROC. OF THE 16TH CONFERENCE ON WORLD WIDE WEB (2007) (describing a family of attacks that can enable an adversary to learn of connections between specific users, in a network); Carter Jernigan & Behram F.T. Mistree, *Gaydar: Facebook Friendships Expose Sexual Orientation*, 14 FIRST MONDAY (2009) (finding that the percentage of a given user's friends who self–identify as gay male is strongly correlated with the sexual orientation of that user).

[39] *See* Kim & Scott, *supra* note 34, at 97 (observing that "Facebook systematically collects large amounts of data about users' activities on the site, such as who their friends are, when they 'like' something, and what links they click").

[40] *See* RadicalXChange Foundation, *The Data Freedom Act,* Working paper 1 (2020), https://www.radicalxchange.org/media/papers/data-freedom-act.pdf [hereinafter *The Data Freedom Act*].

[41] *See* Solon Barocas & Karen Levy, *Privacy Dependencies*, 95 WASH. L. REV. 555 (2020) (arguing that "it can be practically difficult to disentangle whether the information 'belongs' to Alice or to Bob and which of them ought to have control over disclosure decisions"); Viljoen*, supra* note 7, at 580 ("Data production in the digital economy is fundamentally relational.").

[42] *See The Data Freedom Act, supra* note 40, at 1 ("Data about people is always the output of a network of social activity. Even apparently 'individual' data, such as a particular consumer's shopping habits or travel itinerary, is a product of the social world in which that person lives... for example in the context of emails, since any email that is in my inbox inherently exists in somebody else's inbox as well.").

[43] *See* Gergely Biczok & Pern Hui Chia, *Interdependent Privacy: Let Me Share Your Data*, INT'L CONF. FIN. CRYPTOGRAPHY & DATA SECURITY 338 (2013) (describing one user tagging another in a photo as an example of the interdependent nature of data online).

Electronic copy available at: https://ssrn.com/abstract=4105443

disease may have implications not only for the searcher, but also for the searcher's genetic relatives.[44]

In other words, platforms analyze user data not to recognize each individual's uniqueness but to examine how they fit into patterns, clusters, and trends.[45] Solon Barocas and Karen Levy call these relationships *privacy dependencies* and present three categories that describe how the personal attributes of one user are inferred based on their social, physical, or electronic ties with one another.[46] These include tie-based dependencies, similarity-based dependencies, and difference-based dependencies.[47]

The most intuitive way that one person's information can provide details about another is if the second user is captured in the first user's data unintentionally based on their social, physical, or electronic ties, in what Barocas and Levy term *tie-based dependencies*.[48] When Alice uploads a photo from a party she attended, the platform learns about her friend Bob who appears in the photo. Similarly, if Alice uses a virtual assistant or a video-integrated doorbell, the platform may learn information about Bob that is captured even without his knowledge. The platform may also directly prompt Alice to provide information about Bob; for example, when downloading Facebook's Messenger app, users, sometimes unwittingly, give Facebook permission to collect their entire contact list.[49] If enough of Bob's friends join the service, Facebook will be able to construct a web of Bob's social ties even

---

[44] *See* Sylvie Delacroix & Neil D. Lawrence, *Bottom Up Data Trusts: Disturbing the One Size Fits All Approach to Data Governance*, 9 INT'L DATA PRIV. L. 236, 249 (2019) ("Genetic data presents particular challenges because our genome encodes not only information about ourselves but our relatives too: sensitive information can leak through other individuals sharing their genomic data.").
[45] *See The Data Freedom Act, supra* note 40, at 2 (describing the intertwined nature of seemingly personal data); Viljoen, *supra* note 7, at 578, 607.
[46] *See* Barocas & Levy, *supra* note 41.
[47] *See id.*
[48] *See id.*
[49] *See* Chen, *supra* note 32.

Electronic copy available at: https://ssrn.com/abstract=4105443

though he himself has provided no information to the platform and may even prefer to avoid the platform altogether.

In some cases, the disclosure of information by one individual enables an observer to indirectly learn something about another, whether because it highlights a similarity between the two users (in a *similarity-based dependency*) or because it shines a light on the way that one user differs from the others (in a *difference-based dependency*).[50] By analyzing the behavior of an individual and comparing it to patterns of behavior common to many users, platforms are able to make predictions about individual users and infer a broad range of personal attributes that they have not expressly provided.[51] For example, when Alice conducts her shopping on a platform, the platform gains knowledge of her personal attributes as well as her shopping habits. If another user with attributes similar to Alice's was to start shopping on the platform, the platform may offer her some of the same products that Alice purchased. Similarly, if a new user demonstrates similar shopping patterns to Alice, the platform may infer that she shares (some of) Alice's personal attributes.

---

[50] *See* Barocas & Levy, *supra* note 41, at 558, 612; Alessandro Mantelero, *Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection*, 32 COMP. L. & SEC. REV. 238, 239 (2016) (acknowledging the collective dimension of data, in particular in the context of privacy and data protection); Julie E. Cohen, *The Biopolitical Public Domain: the Legal Construction of the Surveillance Economy*, 31 PHILO. TECH. 213, 224 (2018) (acknowledging the importance of pattern detection in platforms' ability to make predictions about their users).

[51] *See* Wachter & Mittelstadt, *supra* note 21, at 506 (describing how platforms can infer data about individuals even if they did not provide it); Solon Barocas & Helen Nissenbaum, *Big Data's End Run around Anonymity and Consent*, in PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT 44, 55 (Julia Lane et al., eds.) (2014) (observing that "insights drawn from big data can furnish additional facts about an individual (in excess of those that reside in the database) without any knowledge of their specific identity or any identifying information").

Electronic copy available at: https://ssrn.com/abstract=4105443

As another example, the detection of patterns across groups serves as the basis for gender classification systems employed by platforms.[52] These systems analyze user data such as pictures, videos, likes, and language patterns, drawing insights from the patterns among users who provided their gender in order to infer the gender of users who did not. Users who did not disclose their gender to the platform but were classified as belonging to a certain gender may feel that their privacy, dignity, and autonomy have been violated, and, in some communities, such inferences may even put individuals at risk of harm, including discrimination and oppression.[53] More generally, by analyzing users' behavior, interests, and social connections, platforms are often able to infer a broad range of personal attributes such as race, sexual orientation, income, political interests, and opinions.[54] Due to platforms' ability to infer user attributes that have not been expressly disclosed, an individual cannot prevent a platform from learning about her by refusing to disclose her data. Effectively, this means that no single individual can decide to withhold her data from platforms.

As Salomé Viljoen argues, the relationships between users who belong to a group enable platforms to use data about one user to infer

---

[52] *See* Yingxiao Wu, Yan Zhuang, Xi Long, Feng Lin & Wenyao Xu, *Human Gender Classification: A Review*, INT'L J. BIOMETRICS 1, 6 (2016) (describing gender classification systems and how they operate); Michal Kosinski, David Stillwell & Thore Graepel, *Private Traits and Attributes are Predictable from Digital Records of Human Behavior*, 110 PROC. NAT'L ACAD. SCI. 5802, 5802 (2013) (demonstrating "that easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender").

[53] *See* Viljoen, *supra* note 7, at 581.

[54] *See* Kristen M. Altenburger & Johan Ugander, *Monophily in Social Networks Introduces Similarity Among Friends-of-Friends*, 2 NATURE HUM. BEHAV. 284, 284 (2018) (finding that "even 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").

Electronic copy available at: https://ssrn.com/abstract=4105443

characteristics of another member of the same group.[55] Furthermore, in order to learn something about a group of people it is enough that a small minority has provided their data. In fact, this is precisely the mechanism that allows researchers to use the results of an experiment conducted on a small number of participants to learn something about the entire population of similar individuals. For example, if a study finds a correlation between participants who rank low on agreeableness and compulsive buying behavior,[56] this finding could enable researchers to use this correlation to infer the buying behavior of individuals who did not participate in the study but whose ranking on the agreeableness scale is known. Yet, despite the substantial effects of inferences based on the relationships between users and the significant role that these relationships have played in the development of the data ecosystem, users' interests as a collective are currently severely underrepresented in regulatory discourse.[57]

In summary, the collective, interdependent nature of personal data means that no single individual can decide on their own how much data they want to disclose to platforms, or what data they want to keep private. Therefore, any intervention in the data ecosystem must be grounded in a deep understanding of the strong collective nature of data and the various dependencies that characterize data. As will be discussed in detail in the following Parts, one substantial policy implication of this finding is the need to generate a collective perspective within the data ecosystem.

---

[55] *See id.* at 578.
[56] *See* Kiran Shehzadi, Muhammad Ahmad-ur-Rehman, Anam Mehmood Cheema & Alishba Ahkam, *Impact of Personality Traits on Compulsive Buying Behavior: Mediating Role of Impulsive Buying*, 9 J. SERV. SCI. & MGM'T. 416 (2016).
[57] *See* Viljoen*, supra* note 7, at 613.

Electronic copy available at: https://ssrn.com/abstract=4105443

II. APPROACHES TO OVERCOMING HARMS FROM INCOMING-VECTOR PERSONALIZATION

Recent sessions of Congress have seen an influx of bills aiming to address the challenges created by incoming-vector personalization.[58] In this Part, we analyze the tools presented in a selection of recent proposals as a reflection more broadly of emerging regulatory approaches to overcoming challenges created by incoming-vector personalization.[59] We begin with a discussion of the challenges of liability and enforcement, including the challenges associated with enforcement of existing anti-discrimination laws with respect to discriminatory effects produced by personalization in the online delivery of certain types of ads, enforcement of data protection laws with respect to various harms resulting from platform personalization, and platform liability in light of Section 230 of the Communications Decency Act and various legislative proposals that would amend the scope of Section 230's protections. We then discuss recent regulatory proposals and analyze the extent to which they incorporate a collective perspective that can enable them to adequately combat the harms they are intended to address. Finally, we present a selection of technological solutions that have been proposed to address the challenges created by incoming-vector personalization.

---

[58] For a detailed review of a wide range of initiatives adopting a proprietarian rationale for regulating data collection and use proposed by technologists, economists, legal scholars, politicians and even a presidential candidate, see Viljoen, *supra* note 7, at 617.

[59] While many of the harms these interventions seek to address are personalization-driven, others stem from non-personalization-related design choices made by platforms. Two examples of the latter category are addictive features and the use of so-called dark patterns to manipulate user behavior, which are the focus, for example, of the DETOUR Act, *supra* note 19, as well as of the SMART Act, *supra* note 19. In this Article, we recognize that personalization of addictive design features or dark patterns can substantially amplify the harms they create. However, the non-personalization-driven aspects of these features are not the main focus of this Article; rather, we limit our focus to the harms arising from incoming-vector content that is personalized for different users based on data collected along the outgoing vector.

Electronic copy available at: https://ssrn.com/abstract=4105443

A. Liability and Enforcement Mechanisms

Enforcement of existing laws may address certain aspects of harmful personalization. U.S. anti-discrimination laws, for instance, prohibit discrimination in ads for housing and job opportunities based on protected attributes such as race, sex, age, religion, and more.[60] In some cases, the content of the ads may not be inherently discriminatory, but the targeting criteria produce discriminatory effects by excluding certain groups on the basis of protected characteristics. For example, Pauline T. Kim and Sharion Scott identified at least three potential ways in which employment recruiting via targeted ad placement on platforms can produce discriminatory effects.[61] The first occurs when advertisers use protected attributes as their targeting criteria, for example, by selecting an audience of only men aged 18 to 40, or by excluding people belonging to an ethnic minority.[62] The second targeting

[60] *See, e.g.,* Fair Housing Act, 42 U.S.C. § 804 (prohibiting discrimination in advertising for housing opportunities); Civil Rights Act of 1964 §§ 703-716, 42 U.S.C. §§ 2000e to 2000e-15 (prohibiting discrimination in job advertisements based on protected characteristics); Age Discrimination in Employment Act of 1967 §§ 2-12, 14-15, 17, 29 U.S.C. §§ 621–634 (prohibiting discrimination in advertising of job opportunities on the basis of age).

[61] *See* Kim & Scott, *supra* note 34, at 98.

[62] In 2016, ProPublica reported on how the Facebook ad targeting platform allows advertisers to place housing ads that explicitly exclude from their targeting criteria users with African-American, Asian-American or Hispanic affinity. *See* Julia Angwin & Terry Parris, Jr., *Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA, Oct. 28, 2016, https://www.propublica.org/article/facebook-letsadvertisers-exclude-users-by-race.
Lawsuits have also alleged that the Facebook ad platform enables the placement of discriminatory advertising, *See, e.g., Bradley v. T-Mobile US*, 2020 U.S. Dist. LEXIS 44102 (N.D. Cal. Mar. 13, 2020) (dismissing a class action lawsuit against T-Mobile and Amazon for allegedly routinely using ad targeting criteria that exclude users over the age of 40 from being presented with job ads they placed on Facebook, but outlining criteria for the plaintiffs to file a new complaint and allowing for additional discovery). In response to the reports of discrimination, Facebook announced changes to its targeting mechanism in order to comply with existing anti-discrimination laws. *See* Julia Angwin, *Facebook Says it Will Stop Allowing Some Advertisers to Exclude Users by Race*, PROPUBLICA, Nov. 11, 2016, https://www.propublica.org/article/facebook-to-stop-allowing-some-advertisers-to-exclude-users-by-race. In 2017, ProPublica found that Facebook still enabled discriminatory targeting of housing ads. *See* Julia Angwin, Ariana Tobin & Madeleine Varner, *Facebook (Still)*

Electronic copy available at: https://ssrn.com/abstract=4105443

mechanism involves an advertiser selecting targeting criteria based on seemingly mundane attributes, such as ZIP code or expressed interests, which are strongly correlated with, and in effect serve as a proxy for, a protected attribute.[63] While such a method of targeting may result in discriminatory effects, it may be difficult to anticipate *ex-ante*.[64] The third way in which targeted job ads can be discriminatory is if the advertiser uses a tool like Facebook's "lookalike audience" feature, which can identify a relevant audience based on a sample group, such as the employer's current workforce.[65] If the sample group is biased, this tool will produce an audience that reflects the same bias.[66]

A strong argument can be made that ad targeting of the first type is prohibited by laws such as Title VII of the Civil Rights Act of 1964 and the Age Discrimination in Employment Act of 1967.[67] However, proving that targeting of the second or third type was unlawfully discriminatory is likely to be difficult and to require additional proof of disparate impact, such as statistical evidence gleaned from workflow data.[68] Further, detection of unlawful instances of ad targeting is limited by a lack of transparency into

---

Letting Housing Advertisers Exclude Users by Race, PROPUBLICA, Nov. 21, 2017, https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin.

[63] *See* Kim & Scott, *supra* note 34, at 98.

[64] For example, in areas with a high degree of residential segregation, location, particularly ZIP code, may serve as a proxy for race. *See* Kim & Scott, *supra* note 34, at 98. In other cases, the demographic characteristics of the audience created by the selected combination of targeting criteria may be more difficult to predict. *See id.* at 98-99.

[65] *See* Meta for Business, About Lookalike Audiences, https://www.facebook.com/business/help/164749007013531 (last visited Jan. 17, 2022).

[66] *See* Speicher et al., *supra* note 31, at 11. Targeting potential employees based on a "lookalike" audience criteria, could also be seen as similar to recruiting via word of mouth. In *Thomas v. Washington Cty. Sch. Bd.*, 915 F.2d 922, 925 (4th Cir. 1990), the court found that advertising for job applicants using existing employees' word of mouth had a discriminatory effect and "serve[s] to freeze the effects of past discrimination," whether the employer had discriminatory intent or not.

[67] *See Kim & Scott, supra* note 34, at 113.

[68] *See id*. at 116.

Electronic copy available at: https://ssrn.com/abstract=4105443

platforms' incoming and outgoing vector flows of information, which regulators and watchdog groups would need in order to investigate possible evidence of discriminatory effects.

The U.S. Federal Trade Commission (FTC), exercising its authority to bring enforcement actions against companies that engage in unfair and deceptive trade practices as well as its investigatory power,[69] has also been active in addressing incoming-vector harms.[70] For example, in 2019, the FTC brought an enforcement action against Devumi, a business that sold fake social media followers, views, and likes to buyers seeking to inflate their influence metrics on platforms—a practice that can facilitate the spread of fake product reviews, spam, manipulation, and disinformation.[71] In its complaint, the FTC alleged that Devumi violated the FTC Act by enabling its customers to mislead the public, thereby providing them with the "means and instrumentalities" to commit deceptive acts or practices.[72] Additionally, in December 2020, the FTC launched an investigation, issuing orders to nine social media platforms to provide information about how they determine which content is shown to which consumers, and the effects of their practices on children and teenagers, among other questions.[73] It has been reported that

[69] *See* 15 U.S.C. § 45(a)(1) (providing that "unfair or deceptive acts or practices in or affecting commerce . . . are . . . declared unlawful"); § 46(b) (providing the Commission with the authority to require certain entities engaged in commerce to file "annual or special . . . reports or answers in writing to specific questions").

[70] *See* Federal Trade Commission, Social Media Bots and Deceptive Advertising, Report to Congress (July 16, 2020).

[71] *See* Complaint for Permanent Injunction and Other Equitable Relief, FTC v. Devumi, LLC, No. 9:19cv81419 (S.D. Fla. Oct. 18, 2019); Stipulated Order for Permanent Injunction and Monetary Judgment, FTC v. Devumi, LLC, No. 9:19cv81419 (S.D. Fla. Oct. 18, 2019).

[72] Complaint for Permanent Injunction and Other Equitable Relief, FTC v. Devumi, LLC, No. 9:19cv81419 (S.D. Fla. Oct. 18, 2019). The court order settling this complaint imposed a $2.5 million judgment against Devumi's owner. *See id.*

[73] *See* Federal Trade Commission, *FTC Issues Orders to Nine Social Media and Video Streaming Services Seeking Data About How They Collect, Use, and Present Information*, Press release, Dec. 14, 2020, https://www.ftc.gov/news-events/press-releases/2020/12/ftc-issues-orders-nine-social-media-video-streaming-services.

Electronic copy available at: https://ssrn.com/abstract=4105443

FTC staffers are currently exploring whether Facebook engaged in deceptive or unfair trade practices in light of whistleblower Frances Haugen's statements in September 2021 regarding the company's internal research showing knowledge of harms resulting from its personalization algorithms.[74]

An additional mechanism for addressing certain incoming-vector harms is the enforcement of existing data protection regulations, such as the GDPR.[75] Personalization may be seen to encroach on privacy rights,[76] for example by undermining individuals' right to be left alone,[77] or undermining individuals' right to take a meaningful part in their self-determination, as well

[74] *See* John D. McKinnon & Brent Kendall, *Federal Trade Commission Scrutinizing Facebook Disclosures*, WALL ST. J., Oct. 27, 2021, https://www.wsj.com/articles/facebook-ftc-privacy-kids-11635289993.

[75] GDPR, *supra* note 19. *See* European Data Protection Board, Guidelines 8/2020 on the targeting of social media users 5 (2020) ("Targeting of social media users may involve uses of personal data that go against or beyond individuals' reasonable expectations and thereby infringes applicable data protection principles and rules."); European Data Protection Supervisor, EDPS Opinion 3/2018 on online manipulation 15 (2018) ("The concern of using data from profiles for different purposes through algorithms is that the data loses its original context. Repurposing of data is likely to affect a person's informational self-determination, further reduce the control of data subjects over their data, thus affecting the trust in digital environments and services."). Other jurisdictions have also recently enacted data protection regulations influenced by the GDPR, such as the Lei Geral de Proteção de Dados (LGPD) in Brazil, the proposed Digital Charter Implementation Act in Canada, and the California Consumer Privacy Act (CCPA) as amended by the California Privacy Rights Act. We discuss the rights provided by the GDPR as a reflection of general regulatory trends with respect to data protection.

[76] *See* Daniel J. Solove, *"I've Got Nothing to Hide" and Other Misunderstandings of Privacy*, 44 SAN DIEGO L. REV. 745, 757 (2007) (including data-driven harms under the umbrella of "privacy" such as "problems of information processing… [that] frustrate the individual by creating a sense of helplessness and powerlessness, but they also affect social structure by altering the kind of relationships people have with the institutions that make important decisions about their lives").

[77] *See* Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 195 (1890); Daniel J. Solove, *Conceptualizing Privacy*, 90 CAL. L. REV. 1088, 1105 (2002) (characterizing the right to be left alone as capturing a common understanding of privacy); Bart van der Sloot, *The Right to be Let Alone: Narrative and Identity in a Data-Driven Environment*, 13 L. INNOV. & TECH. 223, 226 (2021) (proposing a reformulation of "the right to privacy that also includes a right to be protected from information-communication to oneself – a right to be let alone by oneself").

Electronic copy available at: https://ssrn.com/abstract=4105443

as their ability to "maintain relational ties and to develop critical perspectives on the world around them."[78]

Data protection principles, such as data minimization and purpose limitation,[79] likely serve to curb platforms' ingestion, use, and retention of large quantities of fine-grained user data in order to target highly personalized content.[80] Numerous scholars have observed that these principles are incompatible with the big data analytics at the heart of platform personalization because they require platforms to inform their users of the specific nature of future processing of such data which, when using personalization algorithms, may be inherently unforeseeable at the time consent is given.[81] In addition, data controllers and processors must demonstrate an applicable legal basis to justify the processing of personal data, such as consent or legitimate interests,[82] for platform targeting, which may pose challenges particularly in contexts in which profiling and tracking persist across multiple platforms.[83]

---

[78] *See* Julie E. Cohen, *What Privacy is For*, 126 HARV. L. REV. 1904, 1906 (2013). *See also* discussion *infra* Section III.A.

[79] *See, e.g.,* GDPR, *supra* note 19, art. 5(1)(b) (providing that the collection of personal data must be limited to "specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes"); *id.* art. 5(1)(c) (providing that personal data must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed").

[80] *See, e.g.,* Tal Z. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995, 1005 (2017).

[81] *See id.* at 1005-6 (2017); Chris Jay Hoofnagle, Bart van der Sloot & Frederik Zuiderveen Borgesius, *The European Union general data protection regulation: What it is and what it means*, 28 INFO. & COMM. TECH. L. 65, 77-78 (2019) (observing that "[p]urpose limitation strikes at the heart of information-intensive industries, because companies so frequently find utility for data by using and repurposing the data in unforeseeable ways" and that "[i]ndeed, the very purpose of machine learning is to discover patterns not anticipated or even perceivable to people").

[82] GDPR, *supra* note 19, art. 6(1)(a), (f).

[83] *See* European Data Protection Board, Guidelines 8/2020 on the targeting of social media users 16 (noting that the Article 29 Working Party "has previously considered that it would be difficult for controllers to justify using legitimate interests as a legal basis for intrusive profiling and tracking practices for marketing or advertising purposes, for example those that involve tracking individuals across multiple websites, locations, devices, services or data-

Electronic copy available at: https://ssrn.com/abstract=4105443

Further, the processing of special categories of personal data, namely "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation" is prohibited,[84] unless one of a list of delineated exceptions, such as "explicit consent . . . for one or more specified purposes," applies.[85] This creates challenges for data-driven personalization by platforms because the boundaries between these special categories of personal data and other types of personal data are blurred by machine learning algorithms;[86] for example, a user's geolocation information may serve as a proxy for race and ethnicity.

Also in tension with platform personalization is the right not to be subjected to fully automated decision-making processes that substantially impact individuals, unless one of several exceptions, such as explicit consent, is met.[87] With respect to such processing, an individual has the rights to "obtain human intervention," "express his or her point of view," "contest the decision," to know of their existence, and to receive "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[88] Scholars have

---

brokering") (citing Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (2018)).

[84] GDPR, *supra* note 19, art. 9(1).

[85] GDPR, *supra* note 19, art. 9(2)(a).

[86] *See* Zarsky, *supra* note 80, at 1013; European Data Protection Board, Guidelines 8/2020 on the targeting of social media users 5 (2020) ("Recent research suggests that the potential for discriminatory effects exists also without using criteria that are directly linked to special categories of personal data in the sense of Article 9 of the GDPR.") (citing Speicher et al., *supra* note 31).

[87] *See* GDPR, *supra* note 19, art. 22(1)-(2). Also note that the DMA prohibits sharing data between jointly owned platforms, which enables "deep consumer profiling." *See* DMA, *supra* note 110, art. 5(a).

[88] *See id.* arts. 22(3), 13(2)(f), 14(2)(g), 15(1)(h).

Electronic copy available at: https://ssrn.com/abstract=4105443

expressed doubt that these requirements will have a "significant practical impact on automated profiling," but could conceivably apply where "advertising involves blatantly unfair discrimination in the form of web-lining and the discrimination has non-trivial economic consequences," particularly in cases where such consequences occur on a repeated basis.[89]

For decades, platforms have enjoyed legal protection from liability for harmful content posted by their users, under laws such as Section 230 of the Communications Decency Act.[90] Such protection has been both heralded as integral to online free speech and criticized as "an ill-conceived shield for scoundrels."[91] Growing calls to restrict or withdraw the broad protection of Section 230 have inspired bills such as the Protecting Americans from Dangerous Algorithms Act sponsored by Reps. Malinowski (D-N.J.-7) and Eshoo (D-Cal.-18),[92] the Justice Against Malicious Algorithms Act of 2021 sponsored by Rep. Pallone (D-N.J.-6),[93] and the Health Misinformation Act of 2021 sponsored by Sens. Klobuchar (D-Minn.) and Luján (D-N.M.).[94] These proposals would amend Section 230 to remove platforms' protection from liability in certain circumstances, seeking to hold platforms responsible for the active role they take in promoting harmful content to users who are likely to respond to it, thus amplifying its presence and impact.[95] For

---

[89] *See* Isak Mendoza & Lee A. Bygrave, *The Right Not to be Subject to Automated Decision Based on Profiling*, in EU INTERNET LAW 89 (Tatiana-Eleni Synodinou, Philippe Jougleux, Christiana Markou & Thalia Prastitou, eds.) (2017).

[90] 47 U.S.C. § 230 (2018).

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

[92] Protecting Americans from Dangerous Algorithms Act, *supra* note 19.

[93] Justice Against Malicious Algorithms Act of 2021, *supra* note 19.

[94] Health Misinformation Act of 2021, *supra* note 19.

[95] Note that the approaches adopted by the collection of legislative proposals discussed in this Section differ significantly from the proposal in the DSA, which places responsibility for content moderation on the platform, by requiring the largest online platforms to set up a notice-and-action mechanism allowing users to report content they believe is illegal. *See* DSA, *supra* note 19, art. 14. While platforms would not be liable for the content, upon

Electronic copy available at: https://ssrn.com/abstract=4105443

example, the Protecting Americans from Dangerous Algorithms Act would remove immunity from liability for large platforms whose algorithms amplify content involving case involving acts of international terrorism, or interference with civil rights or content neglecting to prevent interference with civil rights under "Reconstruction-era statutes originally designed to reach Ku Klux Klan conspirators" that have been cited as a basis for bringing lawsuits against groups who incited the attacks on the U.S. Capitol on January 6, 2021.[96] The Justice Against Malicious Algorithms Act of 2021 would hold platforms accountable for personalized recommendations that they made knowingly or recklessly and that contributed to physical or severe emotional injury to any person.[97] Finally, the need for the Health Misinformation Act of 2021, which arose during the COVID-19 pandemic, would remove liability protection for platforms whose algorithms promote health misinformation.[98]

B. Enhancing Individual Control via Disclosure and Consent Mechanisms

Many proposals focus on individual control-based approaches,[99] such as notice and consent mechanisms, for addressing harms stemming from

---

receiving notice of the presence of allegedly unlawful content, platforms would be obligated to remove it and notify the poster that it had been removed. *See id.* art. 5(1)(b).

[96] Office of Congressman Tom Malinowski, *Reps. Malinowski and Eshoo Reintroduce Bill to Hold Tech Platforms Accountable for Algorithmic Promotion of Extremism*, Press release (Mar. 24, 2021). *See* Protecting Americans from Dangerous Algorithms Act, *supra* note 19, § 2.

[97] *See* Justice Against Malicious Algorithms Act of 2021, *supra* note 19, § 2(a)(2).

[98] Health Misinformation Act of 2021, *supra* note 19, § 3(a)(1)(B). The Secretary of Health and Human Services would be responsible for determining what content should be considered health misinformation. *See id.* § 3(b).

[99] Enhancement of individual control is one of the rationales underlying the fair information practice principles that have inspired many privacy and data protection regulations, such as the GDPR. *See* Julie E. Cohen, *Turning Privacy Inside Out*, 20 THEORETICAL INQUIRIES. L., 1, 10 (2019) (discussing the challenges raised by privacy as control). Note, however, that some privacy scholars disagree with this framing of the GDPR. *See, e.g.*, Meg Leta Jones & Margot E. Kaminski, *An American's Guide to the GDPR*, 98 DENVER L. REV. 93, 93 (2021)

Electronic copy available at: https://ssrn.com/abstract=4105443

platforms' ability to manipulate users and undermine their autonomy.[100] Such approaches often seek to enhance individual control without recognizing data's collective nature, nor providing meaningful insight into the role that personalization plays.[101] Instead, they often burden individuals with uninterpretable, empty choices, rendering the sense of control they convey a mirage.[102]

Individual autonomy is foundational to modern liberal societies and is a prerequisite for the realization of basic human rights such as freedom of expression, as well as the capacity to shape opinions and values and to choose between right and wrong.[103] The question of whether an action subverts individual autonomy is not always clear-cut; in fact, manipulative behavior

("We endeavor to correct common misconceptions about the GDPR: that it is primarily founded on individual consent (it is not); ... and that it is primarily about individual rights and control (it is equally about risk management and corporate compliance).").

[100] Such questions were raised in Europe in the context of the DSA's disclosure requirement in European Data Protection Supervisor, *Opinion 1/2021 on the Proposal for a Digital Services Act* (2021), at 17 (recognizing that "including information about the recommender system parameters and options in the terms and conditions would only make them difficult to find and understand for data subjects").

[101] *See* Viljoen, *supra* note 7, at 582, 617 ("Individualist theories of informational interests result in legal proposals that … practically fall back on individuals to adjudicate between legitimate and illegitimate information production. This not only leaves certain social information harms unrepresented…"; "[Individualist theories] reduce legal interests in information to individualist claims subject to individualist remedies.").

[102] *See* Woodrow Hartzog, *The Case Against Idealising Control*, 4 EUR. DATA PROT. L. REV. 423, 425 (2018) (detailing the limitations of the privacy as control paradigm); Ella Corren, *A Consent Burden Model: The Failure of Consent in Digital Markets and Elsewhere* (work in progress).

[103] *See* GERALD DWORKIN, THE THEORY AND PRACTICE OF AUTONOMY 10 (1973) ("As a political ideal, autonomy is used as a basis to argue against the design and functioning of political institutions that attempt to impose a set of ends, values, and attitudes upon the citizens of a society."); Susser et al., *supra* note 16, at 14–16 (defining manipulation as "hidden interference that deprives us of authorship over our own choices"); Julie E. Cohen, *Examined Lives; Informational Privacy and the Subject as an Object*, 52 STAN. L. REV. 1426 (2000) (suggesting that autonomy is a prerequisite for participation in the governance of a community); Antoinette Rouvroy & Yves Poullet, *The Right to Informational Self-Determination and the Value of Self-Development: Reassessing the Importance of Privacy for Democracy*, In REINVENTING DATA PROTECTION? 47 (2009) ("Self-determination is an elementary functional condition of a free democratic community based on its citizens' capacity to act and to cooperate.").

Electronic copy available at: https://ssrn.com/abstract=4105443

extends across a spectrum.[104] At one end of the spectrum is mildly manipulative behavior, such as a platform's personalized suggestions to post a "happy birthday" message to a friend's feed or to add another user to one's list of friends. While users may not understand exactly what information such recommendations are based on, they likely recognize this as content created by the platform, and the final decision whether to accept them remains within the user's discretion. At the other end of the spectrum are actions platforms take that users are unaware of, and therefore cannot avoid, such as including users in an experiment attempting to manipulate their mood without notifying them or obtaining their informed consent.[105]

An extensive body of behavioral research calls into question the effectiveness of notice and consent mechanisms.[106] This research demonstrates that individuals often fail to read or understand the implications of platforms' terms of service. Moreover, such agreements are contracts of adhesion, offered on a take-it-or-leave-it basis, precluding the ability of

[104] *See* Tess. M. Wilkinson, *Nudging and Manipulation*, 61 POL. STUD. 341, 342 (2013) (recognizing that there are different levels of manipulation); 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.").

[105] *See* Kramer et al., *supra* note 2 (reporting the Facebook emotional contagion experiment and its outcomes).

[106] *See, e.g.,* Joel R. Reidenberg, et al., *Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding*, 30 BERKELEY TECH. L. J. 39 (2015); Ian Ayres & Alan Schwartz, *The No-Reading Problem in Consumer Contract Law*, 66 STAN. L. REV. 545, 600 (2014); Shmuel I. Becher & Tal Z. Zarsky, *Minding the Gap*, 51 CONN. L. REV., 69, 73 (2019); David A. Hoffman, *Relational Contracts of Adhesion*, 85 U. CHI. L. REV. 1395 (2018); Kevin Litman-Navarro, *We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.,* N. Y. TIMES, June 12, 2019, https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html; Uri Benoliel & Shmuel I. Becher, *The Duty to Read the Unreadable*, 60 BOSTON COL. L. REV. 2255, 2257 (2019); Yannis Bakos, Florencia Marotta-Wurgler & David R. Trossen, *Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts*, 43 J. LEG. STUD. 1, 6 (2014).

Electronic copy available at: https://ssrn.com/abstract=4105443

individual users to negotiate changes to their terms,[107] and individuals, as participants in a knowledge-based economy, lack a meaningful choice to opt out of the use of digital platforms altogether. Therefore, investing the time and effort to read and understand these documents would be inefficient.[108] As a consequence, proposals relying on control-based mechanisms such as notice and consent are burdening individuals with a pseudo-choice that they are not equipped to make and, in the process, absolving platforms of responsibility.[109]

Control-based approaches are reflected in two legislative initiatives recently proposed by the European Commission: the Digital Services Act (DSA) and the Digital Markets Act (DMA).[110] The DSA recognizes that platforms wield tremendous power due to their ability to control the content

---

[107] *See* Clayton P. Gillette, *Rolling Contracts as an Agency Problem*, WIS. L. REV. 679, 680 (2004) (arguing that "failure to read may be perfectly rational, especially given the inability to negotiate around terms").

[108] Research has suggested that if every user read every privacy policy they agreed to in a year, it would result in $781 billion in lost productivity. *See* Aleecia M. McDonald & Lorrie Faith Cranor, T*he Cost of Reading Privacy Policies*, 4 I/S: J. L. & POL'Y FOR INFO. SOC'Y 543, 564 (2008); Melvin A. Eisenberg, *Behavioral Economics and Contract Law* in THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND THE LAW (Eyal Zamir, Doron Teichman eds.) (2014) (noting that "analyzing [the terms of standard form contracts] would often be unduly costly"); Lorrie Faith Cranor, Candice Hoke, Pedro Giovanni Leon & Alyssa Au, *Are They Worth Reading? An In-Depth Analysis of Online Advertising Companies' Privacy Policies*, CONF. ON COMM., INFO. & INTERNET POL'Y (TPRC 2014) (finding a lack of transparency in the privacy policies of 75 online tracking companies and a confusing lack of consistent terminology).

[109] *See* Julie E. Cohen, *How (Not) to Write a Privacy Law*, KNIGHT FIRST AMENDMENT INST. AT COLUM. U. (2021).

[110] *See* DSA*, supra* note 19, art. 29, rec. 62 (providing that "very large online platforms should ensure that recipients are appropriately informed, and can influence the information presented to them"). This approach is consistent with recent trends in EU data protection law as reflected in the GDPR. The DSA aims to bring EU regulation of the data ecosystem up to date and in particular will modernize Directive 2000/31/EC of the Eur. Parl. & Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market 1 (Directive on electronic commerce) (OJ L 178, 17.7.2000); Proposal for a Regulation of the European Parliament and of the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act) [hereinafter DMA]; Caroline Cauffman & Catalina Goanta, *A New Order: The Digital Services Act and Consumer Protection*, EUR. J. OF RISK REG. 758, 760 (2021).

Electronic copy available at: https://ssrn.com/abstract=4105443

that users are presented with and, in tandem, the principle that "with size and power comes responsibility."[111] In line with this approach, the DSA requires very large platforms to provide notice in their terms of service that the content they are viewing has been algorithmically generated and to detail the main parameters used by recommender systems.[112] They must also allow their users the ability to influence the parameters used by recommender systems, including providing at least one option to opt out of recommendations based on profiling.[113] In contrast, the focus of the DMA is the functioning and competitiveness of the market, not the rights of a particular user, and its disclosure mandates are aimed at increasing platform transparency vis-a-vis advertisers,[114] requiring platforms to provide advertisers and publishers with data about the price paid for advertising services.[115]

In contrast to an omnibus legislative proposal like the DSA that seeks to address a wide range of incoming-vector harms, regulatory proposals in the United States tend to focus on combating specific categories of harms, such as those stemming from (1) platform experimentation and (2) the filter bubbles created by platforms' manipulative presentation of personalized content, as we discuss in turn.

---

[111] Eline Chivot, *The New EU Rulebook for Online Platforms: How to Get it Right, Who Will it Impact and What Else is Needed?* 20 EUR. VIEW 121, 124 (2021); *see* DSA, *supra* note 19, art. 29(1), rec. 62.

[112] *See* DSA, *supra* note 19, art. 29. *See also* European Data Protection Supervisor, *Opinion 1/2021 on the Proposal for a Digital Services Act* (2021), https://edps.europa.eu/system/files/2021-02/21-02-10-opinion_on_digital_services_act_en.pdf (suggesting that including information in platforms' terms and conditions is unlikely to enable users to become exposed to them or understand them better, and, instead, "[t]he EDPS strongly recommends to require that such information concerning the role and functioning of recommender systems to be presented separately, in a manner that should be easily accessible, clear for average users and concise").

[113] *See* DSA, *supra* note 19, art. 29(1), rec. 62.

[114] *See* DMA, *supra* note 110, art. 5(g).

[115] *See* Nicolas Petit, *The Proposed Digital Markets Act (DMA): A Legal and Policy Review*, 12 J. EUR. COMPETITION L. & PRAC. 529 (2021).

Electronic copy available at: https://ssrn.com/abstract=4105443

1. Experimentation

The unique position of platforms within the data ecosystem enables them to experiment with the presentation of different types of content and observe how various categories of users respond.[116] Platforms continuously run such experiments, aiming to refine their personalization algorithms, boost the impact of content presented to users, and make ongoing changes to its interface in order to generate increased engagement. However, individuals may not be aware that platforms are experimenting on them, or that the content they are seeing is based on past experimentation that leveraged platforms' unique perspective within the data ecosystem. As one example, Facebook's mood manipulation experiment – studying whether users' emotional state could be influenced by the content they were shown on the platform – sparked widespread criticism from civil society, academics, and regulators alike,[117] prompting Facebook to apologize for how the company

[116] *See* Zeynep Tufecki, *Engineering the Public: Big Data Surveillance and Computational Politics*, FIRST MONDAY (2014), https://firstmonday.org/article/view/4901/4097 (arguing that platforms use computational politics to advance their own interests); Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of Information Civilization*, 30 J. INFO. TECH. 775, 775 (2015) (acknowledging that platforms use experiments to present better personalization); Kramer et al., *supra* note 2 (reporting the Facebook emotional contagion experiment and its outcomes); Evan Salinger & 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) (describing the problematic aspects of the Facebook experiment)**.** Before introducing its new "care" button during the COVID-19 pandemic, Facebook experimented with its use on a subgroup of its users. Several years earlier it conducted a similar experiment for adding a flower reaction before Mother's Day in several markets. While the care button was a success, the flower button was not. Andrew Hutchinson, *Facebook's Testing a New COVID-19-Themed Reaction Emoji*, SOCIALMEDIATODAY, Mar. 31, 2020, https://www.socialmediatoday.com/news/facebooks-testing-a-new-covid-19-themed-reaction-emoji/575152.

[117] *See* Kashmir Hill, *Facebook Manipulated 689,003 Users' Emotions for Science,* FORBES, June 28, 2014, https://www.forbes.com/sites/kashmirhill/2014/06/28/facebook-manipulated-689003-users-emotions-for-science (reporting that Facebook acknowledged the nature of the experiment).

Electronic copy available at: https://ssrn.com/abstract=4105443

had communicated the experiment to the public.[118] Critics claim that Facebook's experiments effectively subverted its users' deliberative capacities,[119] treating them as "tools and fools" and insulting their dignity.[120]

An example of a recent legislative proposal to address the harms of platform experimentation is the Deceptive Experiences To Online Users Reduction (DETOUR) Act. Introduced in 2019 and 2021 by Sens. Mark Warner (D-Va.) and Deb Fischer (R-Nev.) and colleagues, the bill seeks "[t]o prohibit the usage of exploitative and deceptive practices by large online operators."[121] In particular, it obligates platforms that conduct psychological or behavioral experiments on their users to receive users' informed consent and to periodically disclose to users as well as to the general public any experiments being conducted by the platform.[122]

However, because its scope is limited to a narrow subset of personalization in the context of psychological or behavioral experiments, this proposal arguably fails to address harms from other similar types of testing used by platforms to refine their personalization of content to users.

---

[118] *See* Michael Roppolo, *Researcher Apologizes for Facebook Study in Emotional Manipulation,* CBS NEWS, June 30, 2014, https://www.cbsnews.com/news/researcher-apologizes-for-facebook-study-in-emotional-manipulation (reporting on Facebook's apology). Facebook COO Sheryl Sandberg clarified that the company was not apologizing for the experiment itself but rather for the way it was communicated. *See* Gail Sullivan, *Sheryl Sandberg Not Sorry for Facebook Mood Manipulation Study,* WASH. POST, July 3, 2014, https://www.washingtonpost.com/news/morning-mix/wp/2014/07/03/sheryl-sandberg-not-sorry-for-facebook-mood-manipulation-study/ (reporting that she "expressed regret over how the company communicated its 2012 mood manipulation study of 700,000 unwitting users, but she did not apologize for conducting the controversial experiment. It's just what companies do, she said.").

[119] *See* CASS R. SUNSTEIN, THE ETHICS OF INFLUENCE: GOVERNMENT IN THE AGE OF BEHAVIORAL SCIENCE 86 (2016) (explaining that behavior that "subverts the target's rational capacities" can be manipulative).

[120] *See* Wilkinson*, supra* note 104, at 345 ("To manipulate people is to treat them as both tools and fools.").

[121] DETOUR Act, *supra* note 19.

[122] *See id*. The bill also addresses other non-personalization driven harms, in particular certain aspects of addiction by prohibiting design features aimed at cultivating compulsive usage of the platform in children under the age of 13. *See id.* § 3(a)(1)(C).

Electronic copy available at: https://ssrn.com/abstract=4105443

Further, because the bill relies solely on tools enabling each individual to exercise control with respect to the content she sees, it fails to recognize the collective nature of data. For example, if Alice has opted out of platform experimentation, but Bob, a friend of Alice (or someone judged by the platform to be in some way similar to Alice), has not, Alice might still see content Bob has interacted with as part of the experiment, because of the socially-driven nature of many platforms' recommendations.

2. Filter bubbles

Scholars, politicians, and the media have sounded the alarm regarding platforms' role in amplifying extremism and polarization through the targeting of progressively more extreme personalized content to users based on their interests and opinions.[123] Because users are shown content that increasingly reaffirms their existing beliefs and reflects the opinions of users

[123] *See* Julie E. Cohen, *Tailoring Election Regulation: The Platform is the Frame*, 4 GEO. L. & TECH. REV. 641, 647 (2020) (claiming that users are sorted into opposing tribes); Luke Munn, *Angry by Design: Toxic Communication and Technical Architectures*, 7 HUM. AND SOC. SCI. COMMC'N. 1 (2020) ("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. SCI. 1, 5 (2021) (describing how algorithmic decision-making can facilitate and increase polarization, extremism, and inequality); Center for Humane Technology, *A New Agenda for Tech,* VIMEO, Apr. 25, 2019, https://vimeo.com/332532972 (describing the ways in which platforms encourage extremism); Manuel Ricardo Torres-Soriano, *The Dynamics of the Creation, Evolution, and Disappearance of Terrorist Internet Forums*, 7 INT'L J. OF CONFLICT & VIOLENCE 164, 164 (2013) (explaining how online forums help promote radical jihadist positions); Jeff Horowitz & Deepa Seetharaman, *Facebook Executives Shut Down Efforts to Make the Site Less Divisive,* WALL ST. J., May 26, 2020, https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499 (reporting that Facebook acknowledges that its algorithms "exploit the human brains' attraction to divisiveness"). Frances Haugen, Written Statement before the U.S. Senate Committee. on Commerce, Science and Transportation, Sub-Committee on Consumer Protection, Product Safety, and Data Security, 117 Cong. (Oct. 4, 2021) [hereinafter Frances Haugen, Written Testimony] ("The result has been a system that amplifies division, extremism, and polarization — and undermining societies around the world.").

Electronic copy available at: https://ssrn.com/abstract=4105443

similar to them, each newsfeed has the potential to turn into an echo chamber[124] or filter bubble, in which users face little or no exposure to opinions or even news reports that contradict their beliefs.[125] "Interactional polarization" and social fragmentation are vital concerns, [126] as deliberation, persuasion, and compromise with opposing views – central to democratic functions[127] – are precluded by the very nature of the personalized experience each user encounters on online platforms.[128]

---

[124] *See* Dominic Spohr, *Fake News and Ideological Polarization: Filter Bubbles and Selective Exposure on Social Media*, 34 BUS. INFO. REV. 150, 151 (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 lose the inclination to proactively discuss ideas with people or groups of a different opinion.").

[125] *See* Bail, et al., *supra* note 15 at 9216 ("Social media sites are often blamed for exacerbating political polarization by creating "echo chambers" that prevent people from being exposed to information that contradicts their preexisting beliefs."); Guy Aridor, Duarte Goncalves & Shan Sikdar, *Deconstructing the Filter Bubble: User Decision-Making and Recommender Systems*, 14TH ACM CONF. ON RECOMMENDER SYS. 82, 82 (2020) (describing that platforms that offer personalized suggestions can lead users "into filter bubbles where they effectively get isolated from a diversity of viewpoints or content"). Exposure to others teaches individuals about themselves and to shape their opinions. HANNAH ARENDT, THE HUMAN CONDITION 50 (1998) ("The presence of others who see what we see and hear what we hear assures us of the reality of the world and ourselves.").

[126] *See* 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"); Robert M. Bond, et al., *A 61-Million-Person Experiment in Social Media Influence and Political Mobilization*, 489 NATURE 295 (2012) (reporting the results of an experiment showing that Facebook users who were presented with a message encouraging them to vote and information about Facebook friends of theirs who had voted, participated in the election at higher rates than people who were only presented with a message encouraging them to vote, without the social context).

[127] *See* ROBERT HUCKFELDT, PAUL E. JOHNSON & JOHN SPRAGUE, POLITICAL DISAGREEMENT: THE SURVIVAL OF DIVERSE OPINIONS WITHIN COMMUNICATION NETWORKS 1-24 (2004) (explaining that political deliberation between people has the potential to enhance democratic aspects); Diana C. Mutz, *Cross-Cutting Social Networks: Testing Democratic Theory in Practice*, 96 AM. POL. SCI. REV. 111, 111 (2016) ("Political talk is central to most current conceptions of how democracy functions.").

[128] *See* Spohr*, supra* note 124 at 151. *See also* Cohen, supra note 78, at 1907 ("In its ideal form, the liberal self-possesses both abstract liberty rights and the capacity for rational deliberation and choice and is capable of exercising its capacities in ways uninfluenced by cultural context."); Cohen, *supra* note 123, at 659 (discussing the presumption that more

Electronic copy available at: https://ssrn.com/abstract=4105443

It is against this backdrop that the Filter Bubble Transparency Act (FBTA) was introduced in 2019 and 2021 by Sen. John Thune (R-S.D.) and colleagues.[129] The bill seeks to implement disclosure and consent requirements to address the rise of filter bubbles on large platforms.[130] In particular, it requires large platforms to disclose to users that the content they are presented with and the order in which it is presented is determined by an algorithm and based on user-specific data (or inferences based on user-specific data).[131] Additionally, it requires platforms to enable users to opt out of the filter bubble and instead view an input-transparent version of the platform, i.e., a newsfeed that was not algorithmically personalized based on user-provided content.[132] This approach is similar to the DSA's requirement for platforms using recommender systems to notify their users and enable them to opt out of seeing content based on profiling.[133]

---

information will lead people to in depth discourse which is in and of itself a noble goal, but noting that the reality is far from this ideal); Robert Post, *The Constitutional Status of Commercial Speech*, 48 UCLA L. REV. 1, 7 (2000) ("Public discourse is comprised of those processes of communication that must remain open to the participation of citizens if democratic legitimacy is to be maintained."); *see* also FRANK PASQUALE, THE BLACK BOX SOCIETY 61 (2015) ("The power to include, exclude, and rank is the power to ensure which public impressions become permanent and which remain fleeting.").

[129] *See* FBTA, *supra* note 19.

[130] *See* Adi Robertson, *The Senate's Secret Algorithms Bill Doesn't Actually Fight Secret Algorithm*, THE VERGE, Nov. 5, 2019, https://www.theverge.com/2019/11/5/20943634/senate-filter-bubble-transparency-act-algorithm-personalization-targeting-bill.

[131] *See* FBTA, *supra* note 19, § 3(b)(1)(A) ("The person provides notice to users of the platform that the platform uses an opaque algorithm that makes inferences based on user specific data to select the content the user sees.").

[132] *See* FBTA, *supra* note 19, § 3(b)(1)(A) (B) ("The person makes available a version of the platform that uses an input-transparent algorithm and enables users to easily switch between [the two versions].").

[133] *See* DSA, *supra* note 19, art. 29; *see also id.* rec. 62 (requiring that "very large online platforms [] ensure that recipients are appropriately informed, and can influence the information presented to them").

Electronic copy available at: https://ssrn.com/abstract=4105443

One question this bill raises is how an alternative view not based on user-provided content would be generated.[134] For instance, would users of social networks still see content posted, liked, or shared by their social contacts, or content from groups they belong to or pages they have liked (say, in reverse chronological order)? If so, the assumption that any one individual could single-handedly remove herself from the content promoted within the filter bubble while remaining active on platforms reflects a lack of understanding of the collective nature of data – as an individual who opts out would still see a newsfeed laced with polarizing ideas and content that personalization algorithms promoted to her social media contacts.[135]

Without an overhaul of the current approach to control-based mechanisms, it is unlikely such mechanisms will translate into greater protection of individual autonomy.[136] In particular, an effective consent-based mechanism must ensure individuals are able to make meaningful and consequential choices regarding authorized uses of their data, including permissible types of personalization.[137] Additionally, they must be presented

[134] One option discussed in this context is that the default feed would be similar to the sparkle icon option on Twitter. Since 2018, Twitter has provided users with two options to view their newsfeed: either Twitter's choice of top Tweets, or, for those users who opt out of this view by selecting the sparkle icon, tweets from accounts they follow in reverse chronological order. Twitter Support (@TwitterSupport), TWITTER (Dec 19, 2018, 4:39 PM), https://twitter.com/twittersupport/status/1075506037820579841; Will Oremus, *Twitter Has Finally Made It Easy to Set Your Timeline to Reverse-Chronological,* SLATE, Dec. 18, 2018, https://slate.com/technology/2018/12/twitter-reverse-chronological-timeline-setting.html.

[135] *See* Natali Helberger, Max van Drunen, Sanne Vrijenhoek & Judith Möller, *Regulation of News Recommenders in the Digital Services Act: Empowering David against the Very Large Online Goliath*, INTERNET POLICY REVIEW, Feb. 26, 2021, https://policyreview.info/articles/news/regulation-news-recommenders-digital-services-act-empowering-david-against-very-large.

[136] *See* Susan Benesch, *Proposals for Improved Regulation of Harmful Online Content*, Report of the Dangerous Speech Project 23 (2020).

[137] *See* Daniel Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1894 (2013) (recognizing the challenges of a notice and consent regime while expressing concern that regulation compelling certain privacy choices may be too paternalistic). *See also* Viljoen, *supra* note 7, at 594 ("Notice and consent structures the

Electronic copy available at: https://ssrn.com/abstract=4105443

with more than one viable option to choose from, the consent process must not be overly burdensome, and individuals must be meaningfully informed about the ramifications of each choice.[138] Inasmuch as platforms' incentives remain fixed, however, countering the harmful effects of platform personalization will require entrusting a third-party body with a collective perspective as outlined below in Part III.

C. Transparency Mandates

A third category of interventions includes mandates for platforms to disclose certain information regarding personalization to third parties for transparency and accountability purposes. Disclosure requirements that incorporate recognition of the collective aspect of data are a critical component of interventions for overcoming harms such as disinformation and discrimination. The need for statutory mandates for transparency was underscored by recent attempts by platforms to block third parties from collecting information about outgoing- and incoming-vector content. In August 2021, Facebook shut down the accounts of three New York University researchers who were initially granted access to conduct a study regarding political ads on the platform,[139] on the grounds that they had

basic legal relationship between the individual consumer (the data subject) and the digital service provider (the data processor).").

[138] *See* DSA, *supra* note 19, art. 12(1). The DSA seeks to establish a standard for increased clarity for users with regards to the terms and services provided by platforms. The DSA requires platforms to include certain information in "clear and unambiguous language" and "in an accessible format," in policies regarding content moderation as well as information about platforms' use of recommender systems. *See id.* arts. 12, 29.

[139] *See* Laura Edelson & Damon McCoy, *We Research Misinformation on Facebook. It Just Disabled Our Accounts*, N.Y. TIMES, Aug. 10, 2021, https://www.nytimes.com/2021/08/10/opinion/facebook-misinformation.html. This action followed previous efforts by Facebook to thwart third-party transparency tools, including those from ProPublica, Mozilla, and AlgorithmWatch. *See* Jeremy B. Merrill & Ariana Tobin, *Facebook Moves to Block Ad Transparency Tools — Including Ours*, PROPUBLICA, Jan. 28, 2019, https://www.propublica.org/article/facebook-blocks-ad-transparency-tools;

Electronic copy available at: https://ssrn.com/abstract=4105443

violated the platform's terms of service prohibiting the use of automated scraping tools and, in particular, that such scraping posed risks to individual privacy.[140] Facebook has responded to criticism about its lack of transparency by making certain data available to researchers; however, researchers have noted that access has been too limited to enable effective study of harms such as disinformation and manipulation and that federal legislation mandating platform data sharing is urgently needed.[141]

Many proposals promoting transparency require the disclosure of the targeting criteria selected by advertisers and other considerations introduced by the platforms in the actual presentation of ads, which can potentially address harms with respect to discrimination and disinformation.

The Algorithmic Justice and Online Platform Transparency Act,[142] introduced by Sen. Markey (D-Mass.) and Rep. Matsui (D-Cal.-06), seeks to combat platforms' ability to use their algorithms in order to promote content in a discriminatory fashion by mandating transparency.[143] Platforms must retain a record containing data about their algorithmic processes and upon its request, provide the FTC with access to it.[144] The data to be recorded in this

---

See Nicolas Kayser-Bril, *AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook*, ALGORITHMWATCH, Aug. 13, 2021, https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook.

[140] *See* Mike Clark, *Research Cannot Be the Justification for Compromising People's Privacy*, META, Aug. 3, 2021, https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/.

[141] *See* Simon Hegelich, *World view: Facebook needs to share more with researchers*, 579 NATURE 473 (2020); Nathaniel Persily & Joshua A. Tucker, *Report: How to fix social media? Start with independent research*, BROOKINGS, Dec. 1, 2021, https://www.brookings.edu/research/how-to-fix-social-media-start-with-independent-research.

[142] Algorithmic Justice and Online Platform Transparency Act, *supra* note 19.

[143] The bill also promotes tools of disclosure to users; for example, it requires platforms to clearly disclose to users the categories of personal information collected, how it is collected, and what method the platform's algorithms use to promote or withhold content from users. *See id.* § 4(a)(1)(A).

[144] *See id*. § 4(a)(2)(C).

Electronic copy available at: https://ssrn.com/abstract=4105443

database includes information about the personal data collected and how it is used, as well as information about the algorithm, what data was used in its training, and how it was audited to prevent discrimination.[145] If the algorithm promotes ads for services such as housing, education, employment, insurance or credit, the platform must also assess whether the algorithm creates a disparate outcome based on a protected attribute.[146] The bill also requires platforms to publish a publicly available annual report of their content moderation practices.[147]

Disinformation campaigns have interfered in democratic elections and engendered mistrust in democratic institutions and in democracy itself.[148] Such content can incite individuals to harm democratic symbols,[149] commit violent acts, or even participate in genocide.[150] Although disinformation is not an exclusively personalization-driven harm, the harmful effects of disinformation are substantially amplified by platforms' ability to present such content to users who are more susceptible to believing and acting upon

---

[145] *See id*. at § 4(a)(2)(A).
[146] *See id*. § 4(a)(2)(A).
[147] *See id*. § 4(b)(2)(A).
[148] *See* Robert S. Mueller, Report On The Investigation Into Russian Interference In The 2016 Presidential Election, U.S. DEP'T OF JUST. (2019) [hereinafter Mueller Report].
[149] *See* 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.
[150] *See* The World Staff, *In Myanmar, Fake News Spread on Facebook Stokes Ethnic Violence*, THE WORLD, Nov. 1, 2017, https://www.pri.org/stories/2017-11-01/myanmar-fake-news-spread-facebook-stokes-ethnic-violence (describing how fake news posted on Facebook allegedly had a role in facilitating the genocide of Rohingya Muslims in Myanmar); Alexandra Stevenson, *Facebook Admits It Was Used to Incite Violence in Myanmar*, N.Y. TIMES, Nov. 6, 2018, https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html (reporting that Facebook acknowledged it had a certain role in the events). Fake news was also alleged to have incited violent attacks in Sri Lanka in 2018. *See* Amanda Taub & Max Fisher, *Where Countries Are Tinderboxes and Facebook Is a Match*, N.Y. TIMES, Apr. 21, 2018, https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html.

Electronic copy available at: https://ssrn.com/abstract=4105443

it.[151] The spread of disinformation online can also indirectly impact individuals who do not actively participate on digital platforms.[152]

While existing U.S. laws seek to increase transparency by requiring disclosure of the sponsors of political ads on TV, radio, and satellite, such requirements do not apply to ads placed online.[153] Following findings of Russian involvement in the 2016 U.S. presidential election, Sens. Warner (D-Va.), Klobuchar (D-Minn.), and Graham (R-S.C.) introduced the Honest Ads Act in order to uphold the Supreme Court's ruling in *Buckley v. Valeo* that transparency requirements with regards to political ads should "[provide] the electorate with information" and "insure that voters are fully informed" of the identity of who they are listening to.[154] The bill seeks to expand the applicability of the existing disclosure requirements for political ads as set forth in the Federal Election Campaign Act of 1971[155] to online media, thereby requiring platforms to accompany political advertisements with a clear statement disclosing who is financing them.[156] In addition, platforms would be required to maintain a publicly accessible database regarding various details pertaining to political ads placed (or requested to be placed) on them, including, *inter alia*, "a description of the audience targeted by the advertisement."[157] This requirement seeks to establish a collective point of view regarding the ability to detect personalization; however, because it

---

[151] *See* Tomer Shadmy, *Content Traffic Regulation: Addressing Misinformation while Protecting Free Speech* (2022) (under review).
[152] *See* Frances Haughen, Written Testimony, *supra* note 102, at 3 ("Right now, Facebook chooses what information billions of people see, shaping their perception of reality. Even those who don't use Facebook are impacted by the radicalization of people who do. A company with control over our deepest thoughts, feelings and behaviors needs real oversight").
[153] *See* Office of US Senator Mark R. Warner, The Honest Ads Act, Fact sheet, https://www.warner.senate.gov/public/index.cfm/the-honest-ads-act (last updated May 2019).
[154] *Buckley v. Valeo*, 424 U.S. 1, 66, 76 (1976). *See also* Honest Ads Act, *supra* note 19.
[155] 52 U.S.C. § 301.
[156] Honest Ads Act, *supra* note 19, § 5.
[157] *Id.* § 8(j)(2)(B).

Electronic copy available at: https://ssrn.com/abstract=4105443

requires disclosure of only the targeting criteria (as collected along the outgoing vector) and not data about the actual presentation of the content (as presented along the incoming vector), it would not enable a third party to detect correlations between outgoing- and incoming-vector content.

The tools employed by the Honest Ads Act, and their focus on transparency about political *ads* but not other types of content, render it unlikely that the Act will achieve its goal of preventing manipulation of political processes, due to the influence other types of content have on elections. For example, the Mueller report found that much of the disinformation spread online in the period leading up to the 2016 US presidential campaign did not appear in the form of ads.[158] Twitter acknowledged, for instance, that approximately 1.4 million Twitter users had been exposed to content generated by almost four thousand Twitter accounts controlled by the Russian Internet Research Agency (IRA)[159] and consequently spread by unsuspecting Twitter users.[160] Similarly, Facebook profiles were used to promote political rallies and other events organized by the IRA and to invite reporters to attend these events.[161]

Another category of proposed transparency requirements involves the creation of databases for use by researchers. For example, in the European Union, the DSA would require platforms to create a repository of ads presented on their interface that includes a copy of the ad itself, as well as

[158] *See* Mueller Report, *supra* note 148, at 14.
[159] *See id.* at 15.
[160] *See id.* at 25.
[161] *See id.* at 29. Furthermore, public figures and social media influencers may also be involved in spreading political messaging other than political ads, and other types of content may be posted initially for free and then promoted in order to increase the audience size. *See* Anna Reepschlager & Elizabeth Dubois, *New election laws are no match for the Internet*, POLICY OPTIONS, Jan. 2, 2019, https://policyoptions.irpp.org/fr/magazines/january-2019/new-election-laws-no-match-internet.

Electronic copy available at: https://ssrn.com/abstract=4105443

information about the targeting criteria used[162] and aggregate information about the number of users actually presented with the ad (but not information about their personal attributes).[163] The DSA further requires platforms to provide vetted researchers with information that would enable them to identify systemic risks created by platform activity such as dissemination of illegal content or intentional manipulation of platforms' services.[164] Like the Honest Ads Act, because the DSA does not require disclosure of key outgoing vector data about users who saw the ads, it stops short of creating a mechanism that would enable detection of problematic cases of personalized presentation of ads through the identification of correlations between outgoing- and incoming-vector data.

Similarly, in the United States, the Social Media DATA Act,[165] sponsored by Rep. Trahan (D-Mass.-3), would mandate that platforms provide academic researchers and the FTC with access to all ads placed by advertisers,[166] together with details about their targeting and presentation, such as the targeting criteria and mechanism (of the advertiser and the platform) as well as details about the demographics of the audience actually presented with the ads.[167] Another proposal, the Platform Accountability and Transparency Act (PATA), introduced in December 2021 by U.S. Senators Chris Coons (D-Del.), Rob Portman (R-Ohio), and Amy Klobuchar (D-

[162] *See* DSA, *supra* note 19, art. 30(2). *See also* European Commission., Commission Staff Working Document, Assessment of the Code of Practice on Disinformation—Achievements and Areas for Further Improvement SWD (2020), Section 2.1 (stating that the European Democracy Action Plan will also regulate the presentation and transparency requirements of political advertising).
[163] *See* DSA, *supra* note 19, art. 30(2)(e).
[164] *See id.* art. 26(1).
[165] Social Media DATA Act, *supra* note 19.
[166] *See id.* § 2(a)(1)(B). A similar requirement appears in the DSA, *supra* note 19, art. 24.
[167] *See id.* § 2(a)(1)(F). The Algorithmic Justice and Online Platforms Transparency Act, *supra* note 19, § 4(c), also requires that platforms create a library of advertisements including, inter alia, the content of the advertisement, the targeting criteria used and information about the identity of the advertiser and the cost of the advertisement.

Electronic copy available at: https://ssrn.com/abstract=4105443

Minn.), takes a somewhat different approach.[168] It proposes enabling researchers to submit research proposals to the National Science Foundation, and, upon approval, the relevant platforms would be required to provide the data requested.[169] Additionally, the proposal enables the FTC to require ongoing transparency about certain data, even if no particular request has been made by researchers.[170]

In many cases, transparency mandates are paired with safeguards to protect individual privacy when sharing data with third parties. Indeed, privacy is a central concern when platforms are required to disclose user data and is often cited by platforms as a rationale for denying data requests from third parties.[171] PATA includes provisions requiring researchers to submit their research results to the FTC prior to publication in order to ensure final research products do not compromise privacy or other confidential business information.[172]

The Social Media DATA Act envisions that the FTC would establish a working group "tasked with providing guidance on how independent research using social media data can be done in a way that protects academic researcher independence and consumer's rights to privacy,"[173] including consideration of "[u]nder what circumstances privacy preserving techniques such as differential privacy and statistical noise could be used."[174] Differential privacy is a mathematical technique for adding statistical noise

---

[168] PATA, *supra* note 19.
[169] *See id.* § 4.
[170] *See id.* § 12(b)(1).
[171] *See e.g.,* Amanda Holpuch, *Airbnb Refuses to Comply with State Order to Hand Over Users' Data*, THE GUARDIAN, Oct. 8, 2013, https://www.theguardian.com/world/2013/oct/08/airbnb-new-york-users-data.
[172] *See* PATA, *supra* note 19, § 5.
[173] Office of Congresswoman Lori Trahan, Fact Sheet: The Social Media DATA Act of 2021, at 2 (2021), https://trahan.house.gov/uploadedfiles/social_media_data_act_two-pager.pdf.
[174] Social Media DATA Act, *supra* note 19, § 2(c)(4)(C)(ii)(II).

Electronic copy available at: https://ssrn.com/abstract=4105443

to computations to mask the influence of any individuals' data on the outcome,[175] and we agree that it is well suited to the computation of aggregate statistics such as would be needed in order to audit for problematic personalization, as discussed in more detail below in Part III.

Requiring transparency about targeting criteria and information about the actual presentation of ads – in combination with enabling third parties to explore the correlations between personalization along the incoming vector and demographic characteristics revealed along the outgoing vector – could create a broad, collective perspective that would likely enable the detection and study of potential cases of unfair treatment, illegal discrimination, or disinformation. Further, requiring disclosure of ad sponsorship would play a critical role in limiting the ability of malicious parties to spread disinformation.

D. Involvement of External Supervision Mechanisms

Existing law provides various federal agencies with investigatory and enforcement authority with respect to certain incoming-vector harms; for example, the FTC has the authority to demand disclosure of data for the purpose of investigating or bringing enforcement actions against companies

[175] *See* Cynthia Dwork, Frank McSherry, Kobbi Nissim & Adam Smith, *Calibrating Noise to Sensitivity in Private Data Analysis*, THEORY OF CRYPTOGRAPHY CONF. 265 (2006) (introducing the notion of differential privacy); see also Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O'Brien, Thomas Steinke & Salil Vadhan, *Differential Privacy: A Primer for a Non-Technical Audience*, 21 VAND. J. ENT. & TECH. L. 209 (2018) (introducing the notion of differential privacy to a law audience); Ori Heffetz & Katrina Ligett, *Privacy and Data-Based Research*, 28 J. ECON. PERSP. 75, 82 (2014) (explaining the theory and application of differential privacy to a non-technical audience). As another example, the Data Governance Act lists a few privacy preserving techniques that could be used in data sharing: such as anonymisation, pseudonymisation, differential privacy, generalization, or suppression and randomization. *See* Proposal for a Regulation of the European Parliament and of the Council on European Fata Governance (Data Governance Act) COM/2020/767 final [hereinafter DGA], rec. 6.

Electronic copy available at: https://ssrn.com/abstract=4105443

engaging in unfair or deceptive trade practices.[176] Policymakers have introduced a range of proposals for creating additional supervision mechanisms for monitoring or pre-approving certain aspects of platform activity associated with harms to users.[177] For example, the DMA requires platforms to inform the European Commission of their profiling techniques on an annual basis, providing it with ongoing data regarding platforms' profiling practices within the EU on an ongoing basis.[178] A key difference is that enforcement of the DSA is left to each member state, while the European Commission is empowered to enforce the DMA. In addition to providing individual users with the ability to decide whether to participate in experiments, the DETOUR Act seeks to implement another layer of protection in the form of an Independent Review Board (IRB) that would be responsible for approving experiments the platforms want to run. Whereas the notice and consent tools provided to individuals do not take into account the collective nature of data, the IRB may be positioned to incorporate a review of such considerations, including the consequences for groups and society, not only for the individuals directly affected.[179] In another proposal, Susan Benesch suggests creating local independent councils that would set "ethical standards specific to the online distribution of content and cover topics such as terms and conditions, community guidelines, and the content

[176] For a discussion outlining examples of FTC investigations and enforcement actions with respect to incoming-vector harms, see discussion *supra* Section II.A.

[177] For example, the DSA requires member states to establish national Digital Service Coordinators to be in charge of "application and enforcement" of the DSA. *See* DSA, *supra* note 19, art. 38. As part of ensuring these national bodies are in a position to effectively carry out their supervisory role, they are granted broad authority to request access to necessary data from platforms. *See id.*

[178] *See* DMA, *supra* note 110, art. 13.

[179] *See* DETOUR Act, *supra* note 19, § 3(b)(3)(B). The Act does require the board in formulating its rules to "define conduct that does not have the purpose or substantial effect of subverting or impairing user autonomy, decision making or choice … ." *See id.*

Electronic copy available at: https://ssrn.com/abstract=4105443

regulation practices of social media companies."[180] As long as the external supervision mechanism is not motivated by political and other considerations, proposals in this category are likely to be a strong complement to other interventions to address incoming-vector harms.

E. Self-Regulation

Many platforms have adopted approaches to self-regulation with respect to the removal, blocking or restricting of content.[181] At times, they have received criticism for removal of content in certain contentious cases,[182] notably Twitter and Facebook for their decisions to block US President Donald Trump from their platforms.[183] Facebook has implemented a third-

[180] Benesch, *supra* note 136, at 18.

[181] We use the term self-regulation to denote restrictions put in place by platforms themselves, rather than by an external regulator, see Molly Cohen & Arun Sundararajan, *Self-Regulation and Innovation in the Peer-to-Peer Sharing Economy*, 82 U. CHI. L. REV. 116 (2017). While self-regulation could occur at the exclusive initiative of the self-regulating body, it could also be developed in the shadow of the possibility of external regulation. For example, the DSA encourages the European Commission and the European Board for Digital Services (established under Article 47 of the DSA) to develop voluntary industry standards, codes of conduct and crisis protocols to be adopted by platforms as part of their self-regulation. *See* DSA, *supra* note 19, arts. 34-37. Various civil society organizations have also formulated voluntary codes of conduct that platforms and their workers are encouraged to adopt. For example, the Integrity Institute has developed a Code of Conduct and Integrity Institute Oath for platform workers who are part of the Institute's goal to create "an internet that helps individuals, societies and democracies thrive." *See* Integrity Institute, The Integrity Institute Oath, https://integrityinstitute.org/our-values (last visited Jan. 31, 2022). The Oath includes a commitment to put the public first and an acknowledgement that protecting the public is their first job. *See id.*

[182] *See* Kalina Bontcheva, Julia Posetti, Denis Teyssou, Trisha Meyer, Sam Gregory, Claran Hanot & Diana Maynard, *Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression* 147, Int'l Telecomm. Union and U.N. Educational, Scientific and Cultural Organization (2020).

[183] *See* Mike Issac & Sheera Frenkel, *Facebook Says Trump's Ban Will Last at Least 2 Years*, N.Y. TIMES, June 7, 2021, https://www.nytimes.com/2021/06/04/technology/facebook-trump-ban.html.

Electronic copy available at: https://ssrn.com/abstract=4105443

party fact-checking program[184] that aims to limit the spread of disinformation by identifying the source and reviewing the content of posts suspected of being disinformation. In addition, Facebook established an oversight board entrusted with the authority to make binding decisions about what content Facebook should remove from its platform.[185] There are also examples where platforms have aimed to address incoming-vector harms by introducing new user-facing design features; for example, Instagram recently announced a new tool to encourage its users to "Take a Break," in an effort to address criticisms that the platform is intentionally designed to be addictive.[186]

In 2018, the platforms such as Facebook, Google, Twitter and Mozilla signed a Code of Practice on Disinformation. They were later joined by Microsoft and TikTok, as well as advertisers.[187] Representing the first time that platforms and advertisers agreed to adhere to self-regulatory standards to fight disinformation online,[188] the Code recognizes the harm caused by amplification of disinformation, and seeks to strike a balance between individuals' freedom of expression on one hand and the potential harms created by disinformation on the other.[189] The signatories commit to dilute "the visibility of disinformation,"[190] by providing users with tools empowering them to customize their own content, discover content, and "find

[184] *See* Meta Journalism Project, *Meta's Third-Party Fact-Checking Program*, META https://www.facebook.com/journalismproject/programs/third-party-fact-checking (last accessed Jan 22, 2022).
[185] *See* Oversight Board, *Ensuring Respect for Free Expression, through Independent Judgment*, https://www.oversightboard.com (last accessed Jan 22, 2022).
[186] Andrew Hutchinson, *Instagram Tests New 'Take A Break' Feature to Encourage Users to Limit Time in the App,* SOCIALMEDIATODAY, Nov. 10, 2021, https://www.socialmediatoday.com/news/instagram-tests-new-take-a-break-feature-to-encourage-users-to-limit-time/609854/.
[187] *See* .
[188] *See id*.
[189] *See id*. at Preamble.
[190] *See id.* § I.

Electronic copy available at: https://ssrn.com/abstract=4105443

diverse perspectives about topics of public interest."[191] In line with the mechanism proposed in the DSA,[192] the Code requires that users be provided with tools to report content they believe to be disinformation,[193] as well as an explanation as to why they have been presented with particular content.[194] It also recognizes technology will be an integral part of overcoming disinformation and requires parties to invest in technological solutions that will enable prioritizing "relevant, authentic and authoritative information."[195]

Following suit in July 2019, the Australian government published a report offering 23 recommendations "to promote competition, enhance consumer protection and support a sustainable Australian media landscape in the digital age,"[196] including recommendations related to encouraging platforms to develop a voluntary code of conduct on disinformation.[197] Signatories of the Australian Code of Practice on Disinformation and Misinformation, which was developed by a non-profit industry association upon these recommendations, commit to two central requirements.[198] The first includes committing to the code's main objective of providing "safeguards against harms that may be caused by disinformation and misinformation."[199] These include developing tools that aim to reduce the amplification and recommendation of misinformation and disinformation,[200] informing users what types of behaviors are prohibited under the code,[201] and

---

[191] *See id*. § II.D.
[192] *See* DSA, *supra* note 19, art. 14.
[193] *See* European Commission, *supra* note 187, § II.D.
[194] *See id*.
[195] See i*d*.
[196] Australian Government, *Regulating in the Digital Age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry* (2019), at Foreword.
[197] *See* i*d.* at Recommendations 14-15.
[198] *See* DIGI, Australian Code of Practice on Disinformation and Misinformation (2021).
[199] *Id*. art. 1.4.
[200] *See id*. at Outcome 1a.
[201] *See id*. at Outcome 1b.

Electronic copy available at: https://ssrn.com/abstract=4105443

setting up a mechanism allowing users to report misinformation and disinformation.[202] Second, signatories commit to periodically submitting reports describing the signatory's progress towards achieving the Code's goal.[203] The Code has been adopted by leading platforms such as Apple, Facebook, Google, Microsoft, TikTok and Twitter.[204] In an attempt to encourage self-regulation, the EU Artificial Intelligence Act encourages EU member states to develop voluntary codes of conduct that broaden the application of the safeguards created by the AI Act beyond what it strictly requires.[205]

The first-year assessment of the European Code of Practice on Disinformation found that it served as an important basis for dialogue between stakeholders and provided transparency into platforms' policies on disinformation.[206] However, it recognized that a substantial shortcoming of the Code involves a lack of access to platform data, preventing third parties from assessing platforms' adherence to the Code.[207] The reports submitted by signatories to the Australian Code in May 2021, largely reported that the platforms' policy framework was aligned with the Code's requirements.[208]

While platforms' self-regulatory efforts may be a complementary step in the right direction, particularly because they may foster the development of standards which are more aligned with current technological practices,

---

[202] *See id*. at Outcome 1c.
[203] *See id*. art. 5.13.
[204] *See* DIGI, Disinformation Code - About the Code, https://digi.org.au/disinformation-code (last visited Jan. 30, 2022).
[205] *See* Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts COM/2021/206, art. 69.
[206] *See* Eur. Comm'n, Commission Staff working Document, Assessment of the Code of Practice on Disinformation - Achievements and Areas for Further Improvement, SWD (2020), at Section 3.1.
[207] *See id.* at 19.
[208] *See* DIGI, Transparency Reports, https://digi.org.au/disinformation-code/transparency (last visited Jan. 30, 2022).

Electronic copy available at: https://ssrn.com/abstract=4105443

such efforts involve very little external oversight,[209] and have been criticized as "little more than a symbolic activity."[210] Creating a transparency-increasing mechanism, which enables third parties to observe platforms' behavior and take part in tracking their adherence to the standards created by self-regulation, can help alleviate these concerns, as described in more detail below in Part III.

F. Technical Approaches

A number of recent projects looking to address the harms of the data ecosystem take a strongly control-driven perspective, seeking to keep each individual's data in a location controlled by that person, and allowing software under their personal control to dictate whether outside platforms and apps would gain access to their data.[211] While there is both a role and a need for better control of data, we find that such an individualistic perspective misses the nuances of the fundamentally collective nature of data and thus would not be able to meaningfully intervene to prevent incoming-vector harms.

Francis Fukuyama et al. have recently begun to explore a structural intervention they call "middleware," which is software that would enable users to choose the type of content they want to see, how they would like it to be ranked, and the sources they trust to present them with such content.[212]

---

[209] See Julie E. Cohen, *The Regulatory State in the Information Age*, 17 THEO. INQ. L. 369, 401 (2016).

[210] John Braithwaite & Brent Fisse, *Self-Regulation and the Control of Corporate Crime*, in 23 PRIV. POLICING (Clifford D. Shearing & Philip C. Stenning, eds.) 221, 224 (1987).

[211] *See, e.g.,* Solid, *The Solid Project*, https://solidproject.org (last visited Jan. 30, 2022) (explaining that the project enables individuals to "store their data securely in decentralized data stores called Pods… [enabling the individual to] control which people and applications can access it").

[212] *See* Francis Fukuyama, Barak Richman, Ashish Goel, Roberta R. Katz, A. Douglas Melamed & Marietje Schaake, *Middleware for Dominant Digital Platforms: A*

Electronic copy available at: https://ssrn.com/abstract=4105443

Such an intervention is intended to dilute the power that platforms currently have over public and political discourse, and it is technologically situated to minimize friction with the existing ecosystem.[213] However, it is not clear how individual preferences would interact with platform-driven content promotion or personalization in their model. Furthermore, as framed, the middleware proposal does not seek to provide insight into patterns of personalization or their impacts.

A handful of recent technical projects have explicitly aimed to make personalization along the incoming vector more transparent. A number of carefully constructed studies, combining incoming- and outgoing-vector data at a fixed point in time to reveal instances of problematic discriminatory presentation of advertising content by platforms.[214] The limitation of such studies is that, because there is no general infrastructure for collecting such data, they require considerable time and effort to implement, and they are capable of providing visibility into only one isolated issue at one point in time.

Technological Solution to a Threat to Democracy, STAN. U. POL'Y PAPER 2, 6 (2020) ("Middleware's primary benefit is that it dilutes the enormous control that dominant platforms have"); Francis Fukuyama, *Making the Internet Safe for Democracy*, 32 J. DEMOC. 37, 43 (2021) ("[Large platforms] possess not only enormous wealth … but also something of a chokehold over the communications channels that facilitate democratic politics.").

[213] *See supra* Fukuyama, at 43.

[214] *See* sources cited *supra* note 24; Ali et al., *supra* note 17. *See also* Joshua Asplund, Motahhare Eslami, Hari Sundaram, Christian Sandvig & Karrie Karahalios, *Auditing Race and Gender Discrimination in Online Housing Markets*, Proc. Int'l AAAI Conf. on Web and Soc. Media 24, 25 (2020) (demonstrating differential treatment in the presentation of housing ads and property recommendations based on users' race and gender); Lambrecht & Tucker, *supra* note 3, at 2966 (finding that ads promoting job opportunities in the science, technology, engineering, and math fields were presented less often to women, who constitute a prized demographic, and thus a more expensive target-audience for ads. An algorithm that simply optimizes cost effectiveness in ad delivery may deliver ads in an apparently discriminatory way, even if the ads were intended to be gender neutral).

Electronic copy available at: https://ssrn.com/abstract=4105443

Another approach, the Mozilla Rally project,[215] allows individual users of the Mozilla Firefox web browser to sign up, volunteer information about themselves (i.e., outgoing-vector content such as demographic characteristics or answers to surveys), volunteer to allow Mozilla to gather content related to their browsing (such as the URLs of the pages they browse, page content, and how much time they spend on each page), and opt-in to allow pre-approved research projects access to their relevant data.[216] This effort, if widely adopted, could potentially provide broad, meaningful transparency into platform personalization from a collective perspective, due to its access to both (some) incoming and (some) outgoing vector data. One downside is the project's lack of formal privacy guarantees for the potentially quite sensitive data that it gathers. However, the high-level idea is an incredibly promising model.

In summary, we find that many current proposals adopt an individualistic approach. This finding is consistent with many scholars' observations that privacy and data protection have traditionally been conceptualized as individual rights,[217] largely focused on individuals' ability to control the flow of their data through the data ecosystem.[218] This framing of data, however, ignores the current reality in which the process of

---

[215] *See* Mozilla, *It's Your Data. Use It For a Change*, MOZILLARALLY, https://rally.mozilla.org (last visited Jan. 30, 2022).

[216] Mozilla, *Take Control Over Your Data with Rally, A Novel Privacy-First Data Sharing Platform*, DISTILLED, June 25, 2021, https://blog.mozilla.org/en/mozilla/take-control-over-your-data-with-rally-a-novel-privacy-first-data-sharing-platform.

[217] *See* Alessandro Mantelero, *From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era*, in GROUP PRIVACY (Linnet Taylor, Luciano Floridi & Bart van der Sloot, eds.) 2017 (noting that "informational privacy and data protection have been protected as individual rights").

[218] *See, e.g.,* Viljoen, *supra* note 7, at 593. This approach is exemplified by the fair information privacy principles, which have strongly influenced the development of privacy and data protection frameworks in the US and EU and around the world.

Electronic copy available at: https://ssrn.com/abstract=4105443

datafication creates unjust results on a social level,[219] and the fact that it takes analyses of massive amounts of data of many individuals to personalize content in a way that subordinates and manipulates individuals as well as generating collective harms.[220] An individual acting on her own cannot counteract either end of this problem: she alone cannot effectively withhold her data along the outgoing vector and she cannot effectively extricate herself from the harms of incoming-vector personalization.

III. RECOMMENDED DESIGN PRINCIPLES FOR EFFECTIVE INCOMING-VECTOR INTERVENTIONS

In the previous Parts, we described the collective nature of data and discussed how outgoing-vector content provided by one individual can serve (along with the data of many others) to personalize incoming-vector content for other users. We surveyed the central approaches aimed at countering the many harms of incoming-vector personalization, evaluating their strengths and weaknesses using the lens of the collective nature of data, and found that, despite the strengths of certain proposals, the general principles driving many approaches are highly individual-centric.

In this Part, we propose a path forward for addressing the harms of personalization. In particular, we argue the need for a particular form of transparency that we refer to as a *collective perspective*: transparency that allows visibility into correlations between the incoming and outgoing vectors with respect to a large number of people.

One of the central hindrances to the ability of any non-platform actor to overcome the harms of platform personalization is a severe lack of transparency. Without meaningful, effective transparency, we cannot

[219] *See id*. at 617.
[220] *See id.* at 631.

Electronic copy available at: https://ssrn.com/abstract=4105443

properly understand the role that personalization plays in generating or amplifying various harms. At present, there is a lack of clarity regarding even the most basic of questions, such as whether platform personalization contributes to polarization or defuses it.[221] Furthermore, at present it is nearly impossible to detect or measure problematic personalization.

A. What Information Is Needed to Achieve Meaningful, Effective Transparency?

First, meaningful transparency must constitute visibility into the personalized content presented to a *large number* of individuals, not just one or a handful. Indeed, some of the harms that may be induced by incoming-vector personalization are only definable within a broader social context.

For example, if Jane were the only person using a service, it might not be semantically possible for the service to provide Jane with polarizing or discriminatory content, because there would be no other users with whom Jane could be contrasted or compared. More crucially, though, given any definition of what constitutes problematic personalization, the data of only a single person or a small number of people cannot generally be used to determine the presence or extent of the problem. For example, if one wished to show that a particular ad for housing was being displayed in a manner that disproportionately excluded Black individuals, it would not be enough to observe that the ad was shown to a particular White person or was not shown to a particular Black person. Instead, one would need to know the rate of display on a representative sample of the relevant White and Black populations, and one would need enough observations such that measured differences in the rate of display would be statistically significant. Similarly, if one wished to detect content that was being promoted too rapidly or

[221] *See* Bail et al., *supra* note 15.

Electronic copy available at: https://ssrn.com/abstract=4105443

diverging to extremes (due to the risk that such content is often misleading or incendiary),[222] one would need to analyze a broad sample of individuals' incoming-vector content.

The precise number of people's perspectives needed in order to detect problematic personalization depends on the number of types of problematic personalization one wishes to audit for (i.e., when there are more questions to be studied, one must increase the number of observations in order to maintain statistical validity of the conclusions), the sizes of the populations one wishes to study (i.e., if one wishes to detect discrimination against a tiny group, it may be difficult to get enough observations of that group), the prevalence of the problematic phenomenon (i.e., again, one needs many observations in order to see rare events), and the severity of the phenomenon one wishes to detect (i.e., it requires fewer observations to detect extreme discrimination than subtle discrimination). In practice, the actual number of individuals needed to form a useful collective perspective could range from the dozens to the tens or hundreds of thousands.[223]

Second, meaningful transparency must expose patterns and correlations that *relate outgoing-vector content* (such as individual characteristics and actions taken) *to incoming-vector content* at an aggregate level. Visibility into only incoming-vector content could reveal that a certain piece of content was or was not displayed, and how many times, but would be blind to how the decision to present content was *personalized.* It is the individual characteristics and behaviors revealed along the outgoing vector –

---

[222] *See* Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, 359 SCIENCE 1146 (2018) (finding that false news stories spread faster than true ones).

[223] Existing experimental studies, such as those mentioned *supra* note 27 provide some insight into the size of cohorts that have been required to detect specific instances of problematic personalization. *See e.g.* Ali et al., *supra* note 24, for an example, which used a cohort of tens of thousands to hundreds of thousands participants.

Electronic copy available at: https://ssrn.com/abstract=4105443

potentially indicating each individual's age, gender, location, race, religion, political affiliation, income, occupation, medical history, and more – that form the basis of such personalization.[224] Hence, the ability to relate the outgoing vector to the incoming vector is a crucial component of meaningful transparency.[225]

Furthermore, if one wished to detect violations of new rules governing the source of the outgoing vector data that enables personalization – for example, perhaps to only allow personalization on the basis of information that the user explicitly provided, and not on inferences drawn about her[226] – one would need the ability to distinguish the *source of incoming-vector information*.

Insights into incoming-vector personalization must also clearly be *ongoing*, rather than a one-off measure at some point in time, as personalization algorithms and their content (and hence their harms) are constantly changing and evolving. Furthermore, some concerns, such as platforms promoting *increasingly* polarized content, have an inherent longitudinal aspect.

In sum, meaningful transparency thus requires far more than disclosing ad targeting criteria or ad funding details as in the Honest Ads Act,[227] creating databases of ads divorced from the actual outgoing-vector data of those who received them as in the DSA,[228] or focusing primarily on ads as in the Social Media DATA Act.[229] To be effective, transparency with

---

[224] For the definition of *outgoing vector*, see discussion *supra* Section I.A.
[225] Joshua A. Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal & Brendan Nyhan, *Social Media*, *Political Polarization, and Political Disinformation: A Review of the Scientific Literature*, WILLIAM FLORA HEWLETT FOUNDATION, 64 (2018) (reviewing current literature that analyzes the relationship between social media, political polarization and disinformation).
[226] *See* Wachter & Mittelstadt, *supra* note 21, at 610.
[227] Honest Ads Act, *supra* note 19.
[228] DSA, *supra* note 19.
[229] Social Media DATA Act, *supra* note 19.

Electronic copy available at: https://ssrn.com/abstract=4105443

respect to algorithmic personalization must constitute a genuine collective perspective, with ongoing insight into the information provided and observed along the outgoing vector and how it correlates with any personalized content received along the incoming vector, across a large, representative population.

B. What Body Could Be Tasked with Establishing a Collective Perspective?

Currently, platforms are the only actors in the data ecosystem that hold something that approaches a collective point of view that encompasses outgoing- and incoming-vector content of many users. However, past analyses, such as that of Lina Khan and David Pozen, suggest that the incentives of platforms are so misaligned with those of individual users and the public at large that platforms should not and cannot be assigned sole responsibility for detecting, measuring, and mitigating the harms inflected by the personalized content they purvey.[230] It is therefore worth exploring alternative bodies that could be entrusted with the collective perspective.[231]

A body using a collective perspective to detect, measure, or respond to problematic personalization along the incoming vector would need to be *trusted*, in at least two senses. It would need to be trusted to carry out its duties of observation or intervention in the best interest of the individuals and of society, and the analyses it might perform could pertain to quite sensitive

---

[230] Lina M. Kahn & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497 (2017). *See also* Francis Haugen, Written Testimony, *supra* note 123 ("I saw Facebook repeatedly encounter conflicts between its own profits and our safety. Facebook consistently resolves these conflicts in favor of its own profits."); Nathaniel Persily, *Facebook Hides Data Showing It Harms Users. Outside Scholars Need Access*, WASH. POST, Oct. 5, 2021, https://www.washingtonpost.com/outlook/2021/10/05/facebook-research-data-haugen-congress-regulation.

[231] *See* Margot Kaminsky, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1533 (2019) ("Collaborative governance is described, in brief, as a better way to govern fast-changing, risky systems with a high degree of technological complexity").

Electronic copy available at: https://ssrn.com/abstract=4105443

information about some individuals. Such a system should be designed with technological innovations in place so that it is able to make the required measurements (such as quantifying gender disparity in the delivery of a certain type of ad) without any additional party actually needing to access the raw sensitive data of individuals' characteristics or the personalized content that they are shown. In this regard, the Social Media DATA Act[232] is to be lauded for recognizing the potential of modern technology to resolve seeming conflicts between transparency and privacy; often transparency does not require direct access to individuals' sensitive information but just to statistical aggregates that can be computed with small, intentional perturbations in order to provide formal privacy guarantees.

Local differential privacy is one tool that can be used to allow the computation of accurate aggregate statistics (such as the level of correlation between an ad being shown and the race of the viewer) on the basis of personal data to which a large but controlled amount of random noise has been added.[233] Secure multiparty computation tools additionally provide a modern cryptographic toolkit that can remove the need for a monolithic body to be entrusted with correctly and safely carrying out computations related to the presence of problematic personalization.[234] Instead, responsibility for carrying out the duties related to the collective perspective could be shared across a few trusted parties, and as long as most or even some of them were following the protocol, none would be able to corrupt the computation or gain inappropriate access to personal information. Of course, regulatory, and contractual safeguards could provide an additional layer of protection.

---

[232] Social Media DATA Act, *supra* note 19.
[233] *See* sources cited *supra* note 175.
[234] *See* Yehuda Lindell, *Secure Multiparty Computation*, IACR CRYPTOL 300 (2020) (providing an accessible but detailed introduction to the guarantees that secure multiparty computation provides).

Electronic copy available at: https://ssrn.com/abstract=4105443

There are a number of possible models for how the information needed to establish this collective perspective could be sourced by a third party, including potentially by directly intermediating between individuals and platforms, receiving information primarily from individuals, or receiving information primarily from platforms. The governance, funding, and structure of the entity could also take a range of forms, ranging from a government body to a private for-profit or non-profit service heavily regulated by law. In addition, as we detail below, there are a number of possible choices of entities that might be granted access to the insights afforded by the collective perspective.

The proposed EU Data Governance Act (DGA) provides one useful model for establishing trustworthy intermediating bodies.[235] According to the proposal, "data intermediaries" would be required to maintain neutrality, and would not be permitted to use the data for any other purpose other than promoting its lawful exchange.[236] Intermediaries' business model must "assure that there are no misaligned incentives that encourage individuals to make more data available for processing than what is in the individuals' own interest."[237] Furthermore, intermediaries would owe a fiduciary duty to those data holders whose data-sharing they facilitate.[238] The DGA would also recognize data cooperatives, entities that would support users in their data-sharing and serve as a tool in advancing users' ability to make informed and meaningful choices over their data and its sharing, *inter alia,* by enabling "mechanisms to exchange views on data processing" that would best

[235] *See DGA, supra* note 175; *see* Thomas Streinz, *The Future of European Data Law,* in THE EVOLUTION OF EU LAW (Paul Craig & Gráinne de Búrca, eds.) 902, 935 (2021); https://www.consilium.europa.eu/en/press/press-releases/2021/11/30/promoting-data-sharing-presidency-reaches-deal-with-parliament-on-data-governance-act.
[236] DGA, *supra* note 175, art. 11(1).
[237] *Id.* rec. 23.
[238] *See id.* rec. 26.

Electronic copy available at: https://ssrn.com/abstract=4105443

represent members' interests.[239] Such a body could potentially also be positioned to establish the needed collective perspective.

C. How Can Regulation Support the Establishment of the Necessary Collective Perspective?

Legislation must take an active role in establishing or identifying an intermediating body that will establish the collective perspective, tying its hands so that it is worthy of our trust, ensuring that it will have unencumbered access to the information that it needs, establishing mechanisms for the harms that it surfaces to come to light, and providing enforcement mechanisms against those harms.

Regulation should helpfully tie the hands of the intermediating body. It should restrict the body's ability to share any data it receives access to (whether for profit or not) and derivatives of it, and should mandate the use of modern cryptographic and statistical techniques (as discussed above in Section III.B) to minimize the exposure and gathering of sensitive data.

Regulatory intervention will also likely be necessary in order to oblige platforms to cooperate with the monitoring and data collection required in order to establish the collective perspective. This is in line with – although more demanding than – the various transparency mandates currently under discussion as discussed in Section II.C.

Legislation must also support the intermediating body in gaining access to the information it needs. For example, one might consider laws allowing users to install software that enables a third party to collect information about their interaction with a platform, or other legal ability to

[239] *Id*. art. 9 (1)(c).

Electronic copy available at: https://ssrn.com/abstract=4105443

share their data.[240] Such regulation would facilitate direct, non-intermediated access to user data. Currently, platforms restrict users' ability to share content outside the platform in their terms of service and do not allow third parties to scrape content from the platform. Indeed, Facebook has filed lawsuits against individuals and organizations that scraped content from the platforms in violation of its terms of service.[241] Care must be taken to ensure that privacy and security concerns (whether real or fictional) and the corresponding legislation, including the Computer Fraud and Abuse Act,[242] will not be used as an excuse to hamper the effectiveness of the intermediating body.

Legislation would also determine who would have the right to query or access the collective perspective. Under one model, in analogy to the approach taken by PATA,[243] the collective perspective could be made accessible to interrogation by academic researchers, who are subject to oversight by institutional review boards and have applied for and been granted approval to carry out studies on the data. Academic researchers who discovered cases of harmful personalization could share their research findings with the appropriate oversight body in support of potential investigatory and enforcement actions. Alternatively, or additionally, access to the collective perspective could be made available to journalists for investigative reporting purposes. An advantage of either of these first two models is that granting academic researchers and journalists access to the collective perspective opens up the possibility of identifying instances of

[240] A similar proposal appears in the Platform Accountability and Transparency Act, *supra* note 19, § 11.
[241] *See* Jessica Romero, *Taking Legal Action Against Those Who Abuse Our Services*, META, June 18, 2020, https://about.fb.com/news/2020/06/automation-software-lawsuits; Jessica Romero, *Taking Legal Action Against Data Scraping*, META, Oct. 1 2020, https://about.fb.com/news/2020/10/taking-legal-action-against-data-scraping; Jessica Romero, *Combating Scraping by Malicious Browser Extensions*, META, Jan. 14 2021, https://about.fb.com/news/2021/01/combating-scraping-by-malicious-browser-extensions.
[242] 18 U.S.C. § 1030.
[243] PATA, *supra* note 19, § 5.

Electronic copy available at: https://ssrn.com/abstract=4105443

newly-emerging informational harms that are problematic but not considered unlawful under existing law. A third model would involve making the collective perspective directly available to a government agency with investigation and enforcement authority, such as the US Federal Trade Commission, the US Department of Housing and Urban Development, or the US Equal Employment Opportunity Commission (EEOC). In cases where such a body was to identify cases of illegal personalization it could file a complaint, as the EEOC did alleging Facebook facilitated the discriminatory presentation of job ads.[244] Finally, an independent, cooperative entity, such as a data cooperative or data trust, could be established with the explicit purpose of monitoring for unacceptable personalization.

D. What Is the Expected Impact of the Collective Perspective?

The collective perspective, once established, would shed light on the mechanisms by which personalization is contributing to known harms, enable quantification of the severity of harms, and potentially also draw attention to previously unrecognized personalization-driven harms. This would at last provide a basis for informed discourse among academics, policymakers, and society at large, enabling us to grapple with myriad questions such as: how severe is the discrimination in digital advertising of housing opportunities,

[244] In 2019, the Equal Employment Opportunity Commission found that seven employers had violated federal law when advertising jobs on Facebook in a way that excluded women and/or older workers from getting the ads. ACLU, *In Historic Decision on Digital Bias, EEOC Finds Employers Violated Federal Law when they Excluded Women and Older Workers from Facebook Ads,* Press release, Sep. 25, 2019, https://www.aclu.org/press-releases/historic-decision-digital-bias-eeoc-finds-employers-violated-federal-law-when-they (reporting on the decision).
Additionally, The Fair Housing Act, 42 U.S.C. § 804 prohibits discrimination in advertising for housing opportunities. This section served as the basis for the U.S. Department of Housing and Urban Development's charge of discrimination against Facebook in 2019, alleging discrimination in the presentation of ads for housing on the platform, U.S. Dep't Hous. & Urb. Dev., Charge of Discrimination, FHEO No. 01-18-0323-8 (2019).

Electronic copy available at: https://ssrn.com/abstract=4105443

and what role does platform personalization play? Does personalization on the basis of inferred characteristics contribute more to the amplification of misinformation than personalization on the basis of characteristics a user has explicitly provided for the purpose of content-tailoring? How significant is the contribution of algorithmic personalization to the rapid spread of incendiary content?

Once a collective perspective is established, legislation could establish enforcement against problematic personalization – criminal or civil penalties for platforms; flagging, deprioritizing, or blocking of content reflecting problematic personalization. One could also promote adherence to norms regarding personalization by providing key results of telemetry to individual users, regulators, or the public. These norms could be regulatory standards but could also be community norms adopted by individuals who wish to adhere to certain standards even if they go beyond the legal standard. For example, a group of users may not want to see content that has been personalized based on their political position, or may not want to be gender-stereotyped in the personalized content presented to them. Meaningful transparency into algorithmic personalization could give people the power to pressure platforms to live up to their standards.

In summary, we argue that legislators must intervene to establish a collective perspective that will enable us to collectively understand, detect, study, quantify, and respond to problematic personalization. Without such intervention, harmful personalization will continue to harm individuals and society, unchecked, and even largely unobserved.

CONCLUSION

In this Article, we offer an analysis of the structure of the data ecosystem and the incentives that shape it. We identify the importance of and

Electronic copy available at: https://ssrn.com/abstract=4105443

relationship between the outgoing vector (data flowing from a user to a platform) and the incoming vector (content presented to a user by a platform), and offer terminology that enables us to discuss each direction. The terminology provides us not only with words to describe these two data flows, but also allows us to analytically evaluate the various challenges and opportunities presented by each. Surveying the central regulatory and technological approaches aimed at addressing the harms stemming from incoming-vector personalization, we find that a large part of this toolkit will likely be ineffective in its ability to combat incoming-vector harms and we demonstrate that the lack of sufficient recognition of the collective nature of data is a central reason for this failure.

Finally, we offer a path forward involving a radical new level of transparency around platform personalization. In particular, we argue the need for a particular form of transparency that we refer to as a *collective perspective*, affording continuous visibility into correlations between the incoming and outgoing vectors, across a large number of people. We suggest how legislation can support the establishment of a collective perspective.

Electronic copy available at: https://ssrn.com/abstract=4105443