DROWNING IN BIG DATA: ABUNDANCE OF CHOICE, SCARCITY OF ATTENTION AND THE PERSONALIZATION TRAP, A CASE FOR REGULATION

Sofia Grafanaki*

"The Internet is broken."[1]
Walter Isaacson, LinkedIn

"Tim Wu: 'The Internet is the classic story of the party that went sour' "[2]
John Naughton, The Guardian

"The Internet has a dark side. We need a plan for taming it"[3]
Rebecca MacKinnon, World Economic Forum Annual Meeting 2017

"How the Internet Is Loosening Our Grip on the Truth"[4]

* L.L.M., New York University School of Law; M.B.A., Columbia Business School; B.A., Oxford University. Sofia Grafanaki is currently the Chief Operating Officer of Data Elite. Ms. Grafanaki is admitted to practice law in New York, London, and Athens.

[1] Walter Isaacson, *The Internet is broken. Starting from scratch, here's how I'd fix it.*, LINKEDIN (Dec. 14, 2016), https://www.linkedin.com/pulse/internet-broken-starting-from-scratch-heres-how-id-fix-isaacson, https://perma.cc/3B4S-Y3ST (last visited Oct. 5, 2017).

[2] John Naughton, *Tim Wu: 'The Internet is like the classic story of the party that went sour'*, THE GUARDIAN (Jan. 8, 2017, 3:00 AM), https://www.theguardian.com/technology/2017/jan/08/tim-wu-interview-internet-classic-party-went-sour-attention-merchants, https://perma.cc/27HJ-NNDL (last visited Oct. 5, 2017).

[3] Rebecca MacKinnon, *The Internet has a dark side. We need a plan for taming it.*, WORLD ECON. F., (Jan. 16, 2017), https://www.weforum.org/agenda/2017/01/internet-freedom-censorship-regulation, https://perma.cc/S3LB-F232 (last visited Oct. 5, 2017).

[4] Farhad Manjoo, *How the Internet Is Loosening Our Grip on the Truth*, N.Y. TIMES (Nov. 2, 2016), https://www.nytimes.com/2016/11/03/technology/how-the-internet-is-loosening-our-grip-on-the-truth.html?_r=1, https://perma.cc/UKS3-VSYM (last visited Sept. 15, 2017).

Farhad Manjoo, The New York Times

"Today's Internet is Optimized for Noise"[5]
Andy Bromberg, CEO of Sidewire

"The Facebook Algorithm Is Watching You. Here's one way to confuse it."[6]
Adrienne Lafrance, The Atlantic

[1] These eye-catching headlines are just a sample of the headlines that have dominated the news recently. So the secret is out. The Internet can no longer be seen as a utopian place where everyone is equal and freedom is the status quo. Not everything that has to do with technology is simply a "technical" issue. "The technical is political."[7] There are always values and politics involved in the structure, design and architecture of technology; put differently, these issues are socio-technical and have a deep effect on what sorts of citizens we become.[8]

[2] A comprehensive overview of the issues that will shape the Internet's future is conducted by the Winter 2016 issue of *Dædalus, The*

---

[5] Andy Bromberg, *Today's Internet is Optimized for Noise*, MEDIUM: SIDEWIRE (Feb. 10, 2016), https://medium.com/@sidewire/today-s-internet-is-optimized-for-noise-88e8072ff476#.7srircu93, https://perma.cc/X7SN-ZUZL (last visited Oct. 5, 2017).

[6] Adrienne LaFrance, *The Facebook Algorithm Is Watching You: Here's one way to confuse it.*, THE ATLANTIC (Feb. 22, 2017), https://www.theatlantic.com/technology/archive/2017/02/the-algorithm-is-watching-you/517440/?utm_source=atlfb, https://perma.cc/WV3M-4AGD (last visited Oct. 5, 2017).

[7] Yochai Benkler et al., *The Internet*, 145 DAEDALUS, J. AM. ACAD. ARTS & SCI. 5, 6 (Winter 2016).

[8] *See generally* Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373 (2000) (explaining how privacy considerations and design can work together to promote autonomous choice).

Journal of the American Academy of Arts & Sciences, which is devoted to the Internet and curated by Harvard Professor Yochai Benkler and MIT computer scientist and early architect of the Internet David Clarke.[9] One of the key themes that emerge from the issue is that just as the original design choices that engineers faced in the early days of the Internet shaped what the Internet became, i.e. a "'general purpose' network, designed for a variety of uses," so will ongoing and future design choices shape its future.[10] Benkler further explains that the original design choices of the Internet favored "decentralization of power and freedom to act" at the expense of control, and thus maximized entrepreneurial activity and innovation.[11] Market developments have introduced new points of control and will continue to do so; as such, future design choices will inevitably be subject to conflicts of interests between governments, corporate stakeholders and Internet users.[12] The resulting choices will ultimately represent the power dynamics of the environment in which they are made.

[3] For Benkler, power is neither good nor bad, but simply "the capacity of an entity to alter the behaviors, beliefs, outcomes, or configurations of some other entity."[13] What is crucial however, is an effort to continuously identify points of control as they emerge and devise mechanisms that maintain "degrees of freedom" in the network.[14] We are now in the "Age of Big Data," and the technologies and methods that fall

[9] *See* Benkler et al., *supra* note 7.

[10] David D. Clark, *The Contingent Internet*, 145 DAEDALUS J. AM. ACAD. ARTS & SCI. 9, 10 (Winter 2016).

[11] Yochai Benkler, *Degrees of Freedom, Dimensions of Power,* 145 DAEDALUS J. AM. ACAD. ARTS & SCI. 18, 19 (Winter 2016).

[12] *See* Benkler et al*., supra* note 7, at 6.

[13] Benkler, *supra* note 11, at 19.

[14] *See id.* at 20.

under the catchphrase represent one of these "control points."[15] While a precise definition of the term "Big Data" may be elusive, and the uses, tools and techniques associated with big data are wide-ranging, it is helpful to think of the term as reflecting "a paradigm [more] than a particular technology, method, or practice."[16] Viewed this way, "big data […] is a way of thinking about knowledge through data and a framework for supporting decision making, rationalizing action, and guiding practice."[17] As such, Big Data may ultimately allow very few actors to "predict, shape, and 'nudge' the behaviors of hundreds of millions of people."[18] These actors are the few entities large enough to access, control, collect, and analyze vast amounts of data.[19]

[4] In today's information environment, machine-learning algorithms that conduct predictive analytics based on some type of data mining are used in just about every context. They have infiltrated areas such as employment, education, criminal justice, medicine, insurance, retail, media, and culture.[20] At a very broad level such algorithms "learn" from

[15] Steve Lohr, *The Age of Big Data*, N.Y. TIMES (Feb. 12, 2012), http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html, https://perma.cc/SK2U-R7HQ (last visited Oct. 5, 2017).

[16] 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, 46 (Julia Lane et al. eds., Cambridge Univ. Press 2014).

[17] *Id.*

[18] Benkler, *supra* note 11, at 19.

[19] *See id.*

[20] *See* Clay Dillow, *The big data employment boom*, FORTUNE (Sept. 4, 2013), https://www.fortune.com/2013/09/04/the-big-data-employment-boom, https://perma.cc/Z5B9-9NE3 (last visited Oct. 5, 2017); *see also* Doug Wyllie, *How 'Big Data' is helping law enforcement*, POLICEONE (Aug. 20, 2013), https://www.policeone.com/police-products/software/Data-Information-Sharing-Software/articles/6396543-How-Big-Data-is-helping-law-enforcement, https://perma.cc/VK2R-QNNM (last visited Oct. 5, 2017); Bain Insights, *Big Data: Media's Blockbuster Business Tool*, FORBES (Sept. 18, 2014, 1:46 PM),

the past by analyzing it and taking into account what they deem as statistically significant to produce predictions of the future.[21] They are trained to ignore outliers and assume that what has been will be, oftentimes inheriting or creating biases in the process.[22]

[5] While our society is not yet at the point where *all* of our experiences happen in controlled online environments or when we are "connected," the instances and the complexity with which Big Data technologies are involved in our lives are increasing at an unprecedented pace. The speed of technological breakthroughs we are currently experiencing has no historical precedent; for some, we are at the early stages of a "Fourth Industrial Revolution" that is characterized by a convergence of the digital, physical and biological spheres, is evolving at an exponential rather than a linear pace and "will fundamentally alter the way we live, work, and relate to one another."[23]

---

https://www.forbes.com/sites/baininsights/2014/09/18/big-data-medias-blockbuster-business-tool/#68acbf435bf5, https://perma.cc/A6TG-4T34; Bernard Marr, *How Big Data is Changing Insurance Forever*, FORBES (Dec. 16, 2015, 2:28 AM), https://forbes.com/sites/bernardmarr/2015/12/16/how-big-data-is-changing-the-insurance-industry-forever/#71bf5e346e8e, https://perma.cc/ZG8A-B5XG (last visited Oct. 5, 2017); Doug Guthrie, *The Coming Big Data Education Revolution*, U.S. NEWS & WORLD REPORT (Aug. 15, 2013, 3:53 PM), https://www.usnews.com/opinion/articles/2013/08/15/why-big-data-not-moocs-will-revolutionize-education, https://perma.cc/NCM5-BM89 (last visited Oct. 5, 2017).

[21] *See* Will Oremus, *Who Controls Your Facebook Feed*, SLATE (Jan. 3, 2016, 8:02 PM), http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html, https://perma.cc/TUK7-MCNR (last visited Oct. 5, 2017).

[22] *See, e.g.,* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 673-74 (2016) ("Data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that exist in society.").

[23] Klaus Schwab, *The Fourth Industrial Revolution: what it means, how to respond* (Jan. 14, 2016), https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/, https://perma.cc/52PB-5ATZ (last visited Oct. 5, 2017).

[6] Our online experiences are not a simple series of one-to-one relationships with each service we use, but increasingly are more integrated.[24] "Big Data collection and processing, combined with ubiquitous sensing and connectivity, create extremely powerful insights on mass populations, available to relatively few entities."[25] Scholars such as Zeynep Tufekci explain that when "these methods [are] combine[d] with widespread experimentation […], behavioral science that analyzes individuals in a stimulus-response framework and increasingly on-the-fly personalization of platforms, platform companies can nudge users to form beliefs and preferences, follow behaviors, and increase the probability of outcomes with ever-finer precision."[26]

[7] Viewed against this background, today's "design choices" will fundamentally shape the form and structure of our society. The focus of this paper is on the design choices present in the current media environment and more specifically, the social media environment that has created a new type of "platform press."[27] The argument advanced is that as platforms have now become significant distributers of news, the existing design choices that have been made in this context can threaten the viability of a functioning marketplace of ideas as well as the possibility for true choice about receiving valued information.

[8] A recent illustration of the concerns is the issue of fake news in the 2016 presidential election, which brought to the surface a big debate about

---

[24] ELI PARISER, THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU 6–8 (Penguin Press 2011).

[25] Benkler, *supra* note 11, at 23.

[26] *Id.*

[27] Emily Bell et al., *The Platform Press: How Silicon Valley Reengineered Journalism,* TOW CENTER FOR DIG. JOURNALISM 1, 13-16 (2017), https://www.cjr.org/tow_center_reports/platform-press-how-silicon-valley-reengineered-journalism.php, https://perma.cc/RRY9-CM2N.

whether such platforms are in fact media companies, what kind of responsibilities they should bear, the role of section 230 of the Communications Decency Act, and the correct policy approach.

[9] Part I of the paper will provide background information on the current social media environment; the algorithmic filtering that takes place in the curation of news, and the problems that arise from the way it is set up; and will conclude with a call for regulation in the space. Part II will proceed to examine potential objections to regulation in this area, and argue that such objections are not irrefutable. Using examples of regulations that been introduced in somewhat analogous circumstances in the past, Part II will conclude that some types of regulation are constitutionally permissible and can further promote social and constitutional values.

PART I

A. Journalism, media, and social media

[10] Technological advances, and Big Data methods in particular, have fueled a disruptive change in the nature of journalism and the media industry, and have, at a minimum, significantly challenged the position of news organizations as an independent force in a democratic society.[28] A March 2017 report from Columbia's Tow Center for Digital Journalism, provides a comprehensive analysis of the shifts that have taken place in the news media, starting with the widespread availability of broadband which forced traditional publishers to figure out how to transfer their content to a digital environment, and culminated with the emergence of a new kind of "platform press."[29] The shift brought about uncertainty from the beginning, but early on there was a hope that the digital ecosystem would be built on the traditional values and methods of journalism,

[28] *See id.* at 14–16.

[29] *See id.* at 16.

upholding the "core accountability and civic functions of the free press."[30] The rise of interactive journalism, the possibility to comment on articles, and crowdsourcing and citizens journalism[31] while on the one hand represented a decentralization and democratization of the press; also started undercutting the traditional financials of the industry, thereby changing how news creation was funded.[32] Craigslist and Google took over revenues from classifieds and digital subscriptions were hard to generate in an environment where the Internet is thought of as "free." This resulted in, not only a reduction in revenue for news organizations, but more importantly a disruption in the "vertical integration of the industry, which guaranteed access to audiences through privileged and high cost distribution systems."[33]

[11] Even before mobile, social media platforms, and personalized content came along, the nature of journalism started changing. User clicks started playing a big role in the way we experience news, both in the sense of how news items are placed and in how news items are created, the key factor in this context being *popularity*.[34] When a user clicks on a certain news item, the user is then taken to a new page, with different ads. More ads result in more revenue for the site.[35] News items or opinions that get

---

[30] *Id.*

[31] *See generally* Kate Bulkley, *The rise of citizen journalism*, THE GUARDIAN (June 10, 2012, 7:29 PM), https://www.theguardian.com/media/2012/jun/11/rise-of-citizen-journalism, https://perma.cc/Z7VA-78YQ (last visited Oct. 5, 2017) (discussing that currently available technology and distribution platforms allow ordinary men and women to tell the stories that were once the domain of professional journalists).

[32] *See* Bell et al., *supra* note 27, at 15.

[33] *Id.* at 17.

[34] *See id.* at 35.

[35] *See id.* at 47.

clicked on less than others get less visibility, therefore non-mainstream topics and minority opinions are increasingly pushed into the shadows.[36]

[12] To a certain extent there is nothing new here. More "popular" topics have always been placed in the front page of newspapers and the covers of magazines hoping to drive more sales, and media companies have always tried to deliver content that people will consume. They are predominantly for-profit companies with commercial pressures, whose business model has always been partly based on advertisement.[37] Some see this as a "very old problem," based on the fact that the entity that helps deliver the news is not the same as the creator of news, and thus their interests and their understanding of what they should be doing, are not the same.[38]

[13] Even as an old problem however, the precise tracking and analysis that Big Data has enabled, has given it new dimensions. Before, the broadcasting entity, be it print media or TV channel, had some breathing room to use a moral compass in determining what becomes news.[39] While majority tastes arguably played a big role in the past as well, the feedback was not as precise or instantaneous as it in the online environment. In other words, there was room for "editing" in the traditional meaning of the word (as opposed to algorithmic editing), and there was room for the

[36] *See id.* at 36.

[37] *See* Scott Anthony, *What the Media Industry Can Teach Us About Digital Business Models*, HARV. BUS. REV. (June 10, 2015), https://hbr.org/2015/06/what-the-media-industry-can-teach-us-about-digital-business-models, https://perma.cc/PYM9-FL7U (last visited Oct. 5, 2017).

[38] Caroline O'Donovan*, Q&A: Tarleton Gillespie says algorithms may be new, but editorial calculations aren't*, NIEMANLAB (July 8, 2014, 1:56 PM), http://www.niemanlab.org/2014/07/qa-tarleton-gillespie-says-algorithms-may-be-new-but-editorial-calculations-arent, https://perma.cc/43XP-5K53 (last visited Oct. 15, 2017).

[39] *Id.*

editor who feels that it is his responsibility to instigate public debate and deliberate thinking.[40]

[14] There was also room for the journalist that wants to educate the public on a certain subject, and for the outlier that thinks differently. Editors had the flexibility to promote content that would sell in order to fund content that their writers wanted to write about.[41] Companies couldn't know exactly how many eyeballs saw their ad and broadcasting entities didn't know precisely which articles were read, by how many readers and how quickly.[42] We are now living in times where companies know exactly what works and what doesn't, and marketers will not use their budget for ads placed next to content that is not popular (or "clickable").[43] The result is that there is much less room (if at all) for use of a moral compass when deciding what becomes news, as the editor with embedded journalistic ethics may find it increasingly hard to convince shareholders that minority views that bring no advertising revenue are still necessary content. However, this may not be true for all media establishments. Some, like the New York Times, have been very reluctant to bring down the Chinese walls between their newsroom and business department, but this has been to their financial detriment, as a leaked internal newsroom innovation report documented.[44]

---

[40] *See SPJ Code of Ethics: Preamble*, SOC'Y OF PROF'L JOURNALISTS (Sept. 6, 2014, 4:49 PM), http://www.spj.org/ethicscode.asp, https://perma.cc/EU4M-UEAB (last visited Sept. 15, 2017) (stating that "public enlightenment is the forerunner of justice and the foundation of democracy.").

[41] *See* O'Donovan, *supra* note 38.

[42] *Id.*

[43] *Id.*

[44] *See* Myles Tanzer, *Exclusive: New York Times Internal Report Painted Dire Digital Picture*, BUZZFEED (May 15, 2014, 11:06 AM), https://www.buzzfeed.com/mylestanzer/exclusive-times-internal-report-painted-dire-digital-picture?utm_term=.rvEX396lv#.srwla5ZMy, https://perma.cc/D8P6-4B3Y (last visited Oct. 5, 2017); *see also* Jason Abbruzzese, *The Full New York Times Innovation Report*, MASHABLE (May 16, 2014), http://mashable.com/2014/05/16/full-new-york-

[15] Media scholar Chris Anderson points out that another new problem we are facing is a fundamental transformation in journalists' understanding of their audiences.[45] Anderson refers to this kind of journalism that embraces big data as "algorithmic journalism" one that "lacks an emphasis on either *improving* the level of individual knowledge via better information or by filtering out incorrect information."[46] The algorithmic audience he describes has moved away from previous visions of the audience, such as deliberative or agonistic, which correlated to different images of democracy.[47] Rather, it can be quantified and visualized based on algorithms that take into account inputs like search terms, Internet traffic patterns, the advertising market, keyword rates, and the competition.[48] Viewed this way, the problem we are facing is not the classic tension between the entity that delivers the news and the entity that creates the news, but rather a whole new way of creating news.

[16] With the third wave of technological change in the industry, namely the shift to mobile (small screen and smartphones), there was a further disruption illustrated by the move from an "open web" to an

times-innovation-report/#jnM8jITwaPqs, https://perma.cc/P8UQ-5394 (last visited Oct. 5, 2017); Joshua Benton, *The leaked New York Times innovation report is one of the key documents of this media* age, NIEMANLAB (May 15, 2014 5:55 PM), http://www.niemanlab.org/2014/05/the-leaked-new-york-times-innovation-report-is-one-of-the-key-documents-of-this-media-age/, https://perma.cc/4CWE-ZLNM (last visited Oct. 5, 2017).

[45] *See* C.W. Anderson, *Deliberative, Agonistic, and Algorithmic Audiences: Journalism's Vision of Its Public in an Age of Audience Transparency*, 5 INT'L J. COMM. 529, 529 (2011), http://ijoc.org/index.php/ijoc/article/view/884, https://perma.cc/MS6J-V2KT.

[46] *Id.* at 542.

[47] *Id.* at 530.

[48] *See* Daniel Roth, *The Answer Factory: Demand Media and the Fast, Disposable, and Profitable as Hell Media Model*, WIRED (Oct. 19, 2009, 3:00 PM), http://www.wired.com/magazine/2009/10/ff_demandmedia, https://perma.cc/5K3R-53FF (last visited Oct. 5, 2017).

"ecosystem dominated by a small number of platform companies."[49] These new kinds of intermediaries host public expression (i.e. offer storage) and at the same time provide "navigation and delivery of the digital content of others."[50] Communications Professor Tarleton Gillespie has pointed out that the choice of the word "platform" is not accidental, both in their self-characterizations and in the public discourse when describing online content intermediaries; rather, the term reveals the position that such intermediaries are trying to establish.[51]

[17] They have to appeal to users, advertisers, content (media) producers as well as policymakers, and ease the tensions between all these constituencies in a way that implies equality and fairness.[52] Gillespie traces the semantic richness of the term by identifying four distinct categories of the term's etymology: *computational* ("something to build upon and innovate from"), *figurative* (in that the opportunity is both an abstract and a practical promise), *political* ("a place from which to speak and be heard"), and *architectural*.[53] For Gillespie, all four connote to "a 'raised, level surface' designed to facilitate some activity that will subsequently take place."[54] The term is "anticipatory but not causal" and implies an initial "neutrality with regards to activity," and a progressive and egalitarian arrangement that promises to "support those who stand on it."[55] As such, these "platforms" are becoming the primary keepers of the

[49] Bell et al., *supra* note 27, at 17.

[50] *See* Tarleton Gillespie, *The Politics of 'Platforms'*, 12(3) NEW MEDIA & SOC'Y 347, 348 (2010), http://journals.sagepub.com/doi/pdf/10.1177/1461444809342738, https://perma.cc/LYG8-PTPK.

[51] *See id.*

[52] *See id.* at 359–60.

[53] *See id.* at 352.

[54] *See* Gillespie, *supra* note 50, at 350.

[55] *Id.*

cultural discussion as it moves to the Internet, but have managed to position themselves as the antidote to traditional mass media associated with the notion of "elitist gatekeeper[s] with normative and technical restrictions."[56]

[18] Amongst technology platforms, it is social media platforms in particular, such as Facebook, Twitter and Snapchat amongst others, which have added an additional layer of complexity to the already uncertain future of journalism.[57] This is due to their increasing role in how news is circulated and consumed. With significant traffic on news sites coming from social media platforms, editorial decisions may now include calculations addressing the ways in which specific platforms filter content. As audiences move to the mobile and social web, news organizations follow.[58] CNN for instance, outside its core digital outlets (CNN Desktop, CNN Go, CNN mobile web, and CNN apps), uses 5 different video platforms, 12 different social and messaging platforms and 11 emerging and off-platforms.[59]

[19] According to a 2016 survey by Pew Research Center, 62% of U.S. adults get news on social media (the leader of which is Facebook), and 18% do so often.[60] Social media news consumers still get news from a variety of other sources to a fairly consistent degree according to the study, but as compared to a 2013 study, there is a notable increase in news consumption on Facebook, Instagram, and LinkedIn.[61] For some, the

[56] *Id.* at 348, 352.

[57] *See* Bell et al., *supra* note 27, at 9–10, 23–24.

[58] *Id.* at 25.

[59] *Id.* at 27, Figure 2.

[60] *See* Jeffrey Gottfried & Elisa Shearer, *News Use Across Social Media Platforms 2016*, PEW RES. CTR. 2 (May 26, 2016), http://assets.pewresearch.org/wp-content/uploads/sites/13/2016/05/PJ_2016.05.26_social-media-and-news_FINAL-1.pdf, https://perma.cc/KY54-QF6C.

[61] *Id.* at 9.

move of the audience to social media platforms creates additional pressure on reporters to write "click-bait" articles that "'pander to users' worst impulses," as the stories that actually gain traction online are the "too-good-to-check" stories rather than the ones that are comprehensively reported.[62] This concern aligns well with the findings of the Tow Center report, which alarmingly concludes that regardless of whether content is legitimate or illegitimate, "the economics of social platforms incentivize the spread of low-quality content over high-quality material [and] journalism with high civic value […] is discriminated against by a system that favors scale and shareability."[63]

[20] The type of content that platforms prefer is driven by the demands of the advertising market, which generally prefers images and video to text.[64] As a result, platforms try to incentivize publishers to create more video content, which is both more difficult and more costly to produce.[65]

[21] Amongst the other key findings of the Tow Center of Journalism report is the extent of the "influence of social media platforms shapes journalism itself."[66] The report points out that while competition amongst platforms to attract publishers is helping newsrooms reach much larger

---

[62] Timothy B. Lee, *Mark Zuckerberg is in denial about how Facebook is harming our politics*, VOX (Nov. 10, 2016, 10:25 PM), http://www.vox.com/new-money/2016/11/6/13509854/facebook-politics-news-bad, https://perma.cc/5ZXE-T7VL (last visited Oct. 5, 2017).

[63] Bell et al., *supra* note 27, at 10.

[64] *See* Matt Bowman, *Video Marketing: The Future of Content Marketing*, FORBES (Feb. 3, 2017), https://www.forbes.com/sites/forbesagencycouncil/2017/02/03/video-marketing-the-future-of-content-marketing/#d5f22646b535 (last visited Oct. 5, 2017).

[65] *See* Liraz Margalit, *Did Video Kill Text Content Marketing?*, ENTREPRENEUR (Apr. 16, 2005), https://www.entrepreneur.com/article/245003, https://perma.cc/M4JP-CYKL, https://perma.cc/M4JP-CYKL (last visited Oct. 5, 2017).

[66] Bell et al., *supra* note 27, at 10.

audiences, the specific products released and the design standards of each platform are in fact "dictating publisher activity" and making platforms "explicitly editorial."[67] Other studies have also concluded that "press ethics are intertwined with platform design ethics, and press freedom is shared with software designers," who have come to constitute a "liminal press," defined as "people and systems existing outside — but alongside — online news organizations that create the conditions under which mobile news circulates."[68] A high proportion of news content is now designed to be consumed natively, as is the case with Facebook Instant articles, Twitter moments, Apple News, Instagram Stories, and Snapchat discover.[69] In those cases, readers never leave the social media platform, introducing another decision point for publishers which have to determine how to allocate their content and how much native content to "give away" to platforms.[70]

[22] What this translates to is the following: while our experience of everyday news is (at least partly) shaped by the design choices made by online news (aggregator) platforms and social media platforms (which are now replacing publishers and broadcasters as cultural gatekeepers), the methods they use have very different dynamics that we are only just beginning to understand and unpack.[71]

[23] The Tow Center report continues to state that while platforms rely on algorithms to sort and target content, there is a need for "greater

[67] *Id.*

[68] Mike Ananny & Kate Crawford, *Designer or journalist: Who shapes the news you read in your favorite apps?*, NIEMANLAB (Sept. 10, 2014), http://www.niemanlab.org/2014/09/designer-or-journalist-who-shapes-the-news-you-read-in-your-favorite-apps/, https://perma.cc/LD6B-AQX8 (last visited Oct. 5, 2017).

[69] *See* Bell et al., *supra* note 27, at 25–26.

[70] *See id.* at 28.

[71] *See* O'Donovan, *supra* note 38.

transparency and accountability."[72] The audience has no way of knowing how news reaches them and publishers are "at the mercy of the algorithm;" as even though they are producing more content, they do not fully understand who it reaches and how.[73] These algorithms are also constantly tweaked, to optimize the experience and improve performance, and to offset the danger of gaming the system.[74] For example, as companies started realizing Facebook's viral potential, they began gaming the algorithm by creating content that spoke directly to a particular constituency and/or using "clickbait" headlines.[75] This occurred with both legitimate publishers and others who benefited from the spread of misinformation or conspiracy theories.[76]

[24] Further, Robo-Journalism or Automated Journalism is now a reality.[77] The Associated Press [78], Forbes, The Los Angeles Times, and

[72] Bell et al., *supra* note 27, at 10.

[73] *Id.*

[74] *See id.* at 59; *see also* Lee, *supra* note 62.

[75] *See* Bell et al., *supra* note 27, at 62.

[76] *See id.*

[77] *See* Jason Dorrier, *More News Is Being Written By Robots Than You Think*, SINGULARITYHUB (Mar. 25, 2014), http://singularityhub.com/2014/03/25/more-news-is-being-written-by-robots-than-you-think, https://perma.cc/T5LX-UXA4 (last visited Oct. 5, 2017); *see also* Joe Pinsker, *Algorithm-Generated Articles Don't Foretell the End of Journalism*, THE ATLANTIC (June 30, 2014), http://www.theatlantic.com/business/archive/2014/06/algorithm-generated-articles-dont-foretell-the-end-of-journalism/373691/, https://perma.cc/SWM8-R8PZ (last visited Oct. 5, 2017).

[78] *See* Erin Madigan White, *Automated earnings stories multiply*, ASSOCIATED PRESS (Jan. 29, 2015), https://blog.ap.org/announcements/automated-earnings-stories-multiply, https://perma.cc/6HB3-87WA (last visited Oct. 5, 2017) (Associated Press announcing in January 2015 that it was automatically generating more than 3,000 about U.S corporate earnings each quarter).

Thomson Reuters, all use some kind of algorithm that analyzes data and creates news reports.[79] While the technology is still at an early stage and limited to routine stories for repetitive topics (such as sports and weather), it seems to be here to stay. Proponents see a big potential upside for journalists, who will be "free to do more reporting and less data processing" while the robots do all the drudge work, but the algorithms immediately raise concerns about transparency, accountability, and potential implications for society and democracy.[80]

[25] In its Guide to Automated Journalism produced in January 2016, the Tow Center for Digital Journalism observes that little is known about news consumers' demand for algorithmic transparency and the extent to which they want to understand how such algorithms work; but what seems unquestionable is that "automated journalism will substantially increase the amount of available news, which will further increase people's burden to find content that is most relevant to them." [81] The report notes that this will likely increase the importance of search engines and news aggregators, and therefore reemphasize concerns about filter bubbles and potential fragmentation of public opinion.[82] Concluding, the report calls for further research on the potential effects of personalization and the extent to which algorithms can be trusted as a mechanism for providing checks and balances, identifying important issues, establishing a common agenda for the democratic process of public opinion formation, and the

[79] *See* Shelley Podolny, *If an Algorithm Wrote This, How Would You Ever Know?*, N.Y. TIMES (Mar. 7, 2015), http://www.nytimes.com/2015/03/08/opinion/sunday/if-an-algorithm-wrote-this-how-would-you-even-know.html (last visited Oct. 5, 2017).

[80] *Id.*

[81] Andreas Graefe, *Guide to Automated Journalism*, TOW CTR. FOR DIG. JOURNALISM at 12 (Jan. 7, 2016), http://towcenter.org/research/guide-to-automated-journalism/, https://perma.cc/K6QQ-N8N7.

[82] *See id.* at 46.

implications for democracy "if algorithms are to take over journalism's role as a watchdog for government."[83]

B. Personalization and "Relevance"

[26] What I hope the preceding discussion made clear, is that in the current information environment, proliferation of content has placed individuals in a constant challenge of navigating through an overwhelming amount of information. This inevitably calls for some type of information curation (filtering), and while the concept of curation in the media environment is not a new idea, the way such curation is conducted has changed. When individuals consume information online, they increasingly do so via the intervention of an algorithm, trained to select the particular information flow each individual gets, thus playing a significant role in how news gets distributed.[84]

[27] Platforms such as Facebook, Twitter, Google Search, and Blogger, for instance, host information and public expression produced by others and make important decisions about how it gets circulated.[85] Their algorithms, after learning about users from the past, are tasked to (i) deliver content that is "relevant" to each user (i.e. personalized content) and (ii) decide what is of importance (or "trending"), based on data about users' past content consumption.[86] This paper is concerned primarily with the algorithms aimed at the first of these tasks, i.e. personalization algorithms. To the extent that I discuss algorithms that filter what is

[83] *Id.* at 47.

[84] *See id.* at 11.

[85] *See* Tarleton Gillespie, *Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem*, 20 INFO., COMM. & SOC'Y (ISSUE 1) 63, 73-74 (2017), http://dx.doi.org/10.1080/1369118X.2016.1199721, https://perma.cc/58RS-9J98.

[86] *See* Will Oremus, *Who Controls Your Facebook Feed*, SLATE (Jan. 3, 2016, 8:02 PM), http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html, https://perma.cc/RW5S-SEHD (last visited Oct. 5, 2017).

trending or popular, I do so to enrich the conversation and provide context, but I am not suggesting they be subject to the regulation I propose in this paper. I repeatedly use Facebook as an example because its widespread use, and the familiarity that comes with such use, helps to illustrate the discussion. While the examples may be Facebook specific, I do not intend to limit the recommendation of the need for regulation to Facebook alone.

[28] Personalized media content is also not a new idea. MIT Media Lab founder Nicholas Negroponte talked about the "Daily Me" in 1995, describing it as virtual newspaper that would liberate individuals from the choices of different curators of local papers or TV networks, and would instead contain only stories you wanted to see and information you wanted to know.[87] On the other side of this, scholars like Cass Sunstein have warned about the dangers of personalized media content, emphasizing that when individuals live in their own echo chambers it leads to polarization, segregation, and bad policies coupled with inability to agree on good ones.[88] It can also make individuals vulnerable to believe falsehoods. For Sunstein, such an environment is antithetical to the substantive requirements for deliberative democracy, and ultimately drives individuals into a "prison of their own design."[89]

[29] New Big Data technologies have accelerated this personalization process exponentially, by both creating and feeding off a changing society

[87] NICHOLAS NEGROPONTE, BEING DIGITAL 153 (Alfred A. Knopf 1995).

[88] *See generally* CASS SUNSTEIN, REPUBLIC.COM (2001) (discusses how "personalized" news outlets and web sites have generated a society of extremist and closed-minded individuals); CASS SUNSTEIN, REBUBLIC.COM 2.0 (2007) (discussing polarization within the democratic society spurred by the dissemination of information via the internet and other social media forums); CASS SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA (2017) (standing for the proposition that the internet and other platforms of social media are facilitating the demise of democracy) [hereinafter #REPUBLIC].

[89] #REPUBLIC, *supra* note 88, at 12.

that is becoming more exhibitionistic[90] and more intrusive. On the one hand, there is an unprecedented deliberate "sharing" of personal information in the context of social networks, and on the other hand, we are becoming increasingly dependent on the use of apps and the internet of things; which have to track, collect, process, and oftentimes disclose intimate details, about their users in order to be useful.[91] In simplistic terms, this state of affairs enables personalization algorithms to create a version of the "Daily Me" for us.

[30] The personalization process is, by definition, subjective, in that it is based on the perspective of each respective user, viewed individually. This contrasts with algorithms that decide what is "trending," where the decision is based on some objective criteria of what counts as trending in the general user base.[92] Put differently, whereas trending algorithms identify and highlight what is popular with *us* broadly, personalization algorithms identify and highlight what might be relevant to *an individual user* specifically.[93] The key factors for the personalization process are *context* and *relevance.*[94] Context is the factor that gives rise to information privacy concerns, as entities will try to collect as much data as possible

[90] *See* Bill Davidow, *The Internet 'Narcissism Epidemic'*, THE ATLANTIC (Mar. 26, 2013), https://www.theatlantic.com/health/archive/2013/03/the-internet-narcissism-epidemic/274336/, https://perma.cc/6WMA-T72D (last visited Oct. 5, 2017).

[91] *See* Peggy Drexler, *The Problem Isn't Over-Sharing. Its Over-Following*, TIME (Oct. 23, 2014), http://time.com/3535342/oversharing-overfollowing/, https://perma.cc/RL6X-45FP (last visited Oct. 5, 2017).

[92] *See* Sarah Needle, *How Does Twitter Decide What Is Trending?*, RETHINK MEDIA (July 13, 2016), https://rethinkmedia.org/blog/how-does-twitter-decide-what-trending, https://perma.cc/32AY-6EFT (last visited Oct. 5, 2017).

[93] *See* Tarleton Gillespie, *#TrendingisTrending: When Algorithms Become Culture*, forthcoming in *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies* (Robert Seyfert and Jonathan Roberge, eds., 2016) (discussing algorithms as they relate to each user's personal preferences).

[94] *See* Oremus, *supra* note 86.

about an individual and use that data to make inferences that can provide a competitive advantage.[95] The explosion of Big Data technologies translates to an explosion in the range of discoverable contexts, and therefore an increase in the data points that go into the algorithm.[96] The privacy concern is that "an individual may give out bits of information in different contexts and instances, each [one] appearing innocuous," but the aggregation and processing of information can make the resulting picture very invasive of private life.[97] For example, when Facebook combines data about a user's behavior on the site with information from data brokers, they can learn the user's income, net worth, lines of credit, and value of their home; as well as whether the user listens to the radio, or has donated to charity.[98]

[31] Relevance on the other hand, is the force behind filter bubbles and echo chambers, meaning user insulation in a space where they only see content and posts that agree with their preexisting beliefs.[99] Most personalization algorithms are prone to produce such filter bubbles, as their goal is to show users content that will be the most relevant and

[95] *See* Jonathan Salem Baskin, *Privacy Issues Could Threaten the Future of Commercial Social Media*, FORBES (May 28, 2014, 11:08 AM), https://www.forbes.com/sites/jonathansalembaskin/2014/05/28/privacy-issues-could-threaten-the-future-of-commercial-social-media/#76df14286e29, https://perma.cc/A3H5-B4CD (last visited Oct. 5, 2017).

[96] *See id.*

[97] *See* Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy,* 53 STAN. L. REV. 1393, 1452 (2001).

[98] *See* Caitlin Dewey, *98 personal data points that Facebook uses to target ads to you*, WASH. POST (Aug. 19, 2016), https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/?utm_term=.5a8f0d61b5e6, https://perma.cc/295E-T9ZA (last visited Oct. 5, 2017).

[99] *See generally* Pariser, *supra* note 24 (discussing the customization of search results for online users based on personal data).

engaging.[100] Facebook for instance gives each post a relevancy score for each user, measuring whether and how that user is likely to interact with (i.e. click, like, share, and comment on) the specific post.[101] Unsurprisingly, the more the algorithm "knows" about one's past, the better it can predict what this person will want/interact with in the future.[102] This "past" the algorithm needs, is formed from assumptions and inferences based on as much data as possible about an individual, including the individual's characteristics, preferences, habits, personality traits, and what type of content they have engaged with in the past.[103] Based on this past, the algorithm then conducts predictive analysis and determines the choices an individual is given.[104]

[32] In the media/social media environment, these choices are represented by content items that individuals can subsequently click on.[105] In turn, user interaction with these new "choices," translates into new data points for the algorithm.[106] However, a user cannot click on content that the user does not see. The user's choices are already narrowed by the assumptions and inferences made by the algorithm, so when that user actually clicks on a content item, it is inevitable that the initial assumption will be strengthened.[107] This is how self-reinforcing (feedback) loops are

[100] *See* Josh Constine, *How Facebook News Feed Works*, TECHCRUNCH (Sept. 6, 2016), https://techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed/, https://perma.cc/L5VP-GBJY (last visited Oct. 5, 2017).

[101] *See* Oremus, *supra* note 86.

[102] *See id.*

[103] *See id.*

[104] *See id.*

[105] *See id.*

[106] *See* Oremus, *supra* note 86.

[107] *See id*.

created, and how personalization algorithms can ultimately keep us trapped in our comfort zone. These same types of feedback loops are present in other contexts as well, such as offers of credit, where denying individuals an extension credit for bad credit records ends up strengthening existing biases because those individuals are never given the chance to build good credit. In the case of news and social media, psychologists and social scientists have repeatedly illustrated that people tend to choose what feels comfortable and confirms their existing opinion and biases when confronted with diverse information choices.[108] In a social media environment where the content options rise towards infinity, it is not hard to be presented with *a lot* of content that matches one's demographic characteristics and/or political convictions. However, when diverse groups start seeing only points of view matching their characteristics, mutual understanding between groups becomes harder, and, according to social scientists, can lead to "group polarization;". a term which refers to the phenomenon of like-minded groups engaged in deliberation, ending in a strengthening of the original position and a move towards a more extreme point.[109]

[33] An interesting illustration of the filter bubble phenomenon and its consequences was the Wall Street Journal's "Red Feed Blue Feed" feature during the 2016 Presidential Election, which showed two different versions of a newsfeed on the same topic according to ideological affiliations, and demonstrated just how easy it is to be insulated in content that confirms our preexisting beliefs.[110] Interestingly, commentators have pointed out that because the human brain has difficulty distinguishing between "big" and "huge" numbers (meaning that "thousands of people"

[108] *See* Sunstein, *supra* note 88, at 124; *see also* Farhad Manjoo, *How the Internet is Loosening Our Grip On the Truth,* N.Y. TIMES, Nov. 3, 2016, https://www.nytimes.com/2016/11/03/technology/how-the-internet-is-loosening-our-grip-on-the-truth.html?_r=0, https://perma.cc/FLL4-MJR8.

[109] *See* Sunstein, *supra* note 88, at 68.

[110] *See* Jon Keegan, *Blue Feed, Red Feed*, WALL ST. J., May 16, 2016, https://graphics.wsj.com/blue-feed-red-feed, https://perma.cc/6EDN-QTKZ.

and "millions of people" will be processed similarly by the human brain), when users are separated into groups, they can have a false sense that such groups are representative of a majority.[111]

[34] According to the entities using personalization algorithms, the ultimate goal of the effort to deliver personalized content is to serve the consumers better.[112] Assuming the algorithms actually reflect user's choices, Cass Sunstein calls this "consumer sovereignty in action," where individual consumers choose (content in this case) exactly as they wish.[113] Sunstein disputes the notion that a system of communications has as its purpose to ensure that citizens are exposed to only and exactly the content that they want to receive.[114] He distinguishes this type of consumer sovereignty from political sovereignty, the idea on which free nations are based, which stands on different fundamental values.[115] For Sunstein, "political sovereignty embodies democratic self-government, understood as a requirement of 'government by discussion,'" and does *not* account for people as simply having fixed tastes and preferences, there to be discovered.[116] These two concepts can get blurred when the same platform appeals to both our role as consumers, which calls on the concept of consumer sovereignty, and our role as citizens in a democracy, which

[111] *See* Tobias Rose-Stockwell*, How We Broke Democracy*, MEDIUM (Nov. 11, 2016)*,* https://medium.com/@tobiasrose/empathy-to-democracy-b7f04ab57eee, https://perma.cc/2Q69-CUXM.

[112] *See*, *e.g.*, Eugene Kim, *Mark Zuckerberg Wants To Build The 'Perfect Personalized Newspaper' For Every Person In The World*, BUSINESS INSIDER (Nov. 6, 2014), http://www.businessinsider.com/mark-zuckerberg-wants-to-build-a-perfect-personalized-newspaper-2014-11, https://perma.cc/57SN-7AZG (during an interview with Facebook founder Mark Zuckerberg, he says his goal is "to build the perfect personalized newspaper.").

[113] #REPUBLIC, *supra* note 88, at 53.

[114] *See id*. at 54.

[115] *See id.*

[116] *Id.*

implicates political sovereignty. As consumers, we may be better served by free markets; but "if free consumer choices result in insufficient understanding of public problems," we cannot rely on free markets in our capacity as citizens of a democratic society.[117] As citizens we need more than just access to information; we need access to informed debate, which presupposes a certain diversity of viewpoints to begin with.

[35] Moreover, the *relevance* that companies are providing has a built-in assumption that individuals will follow the same behavioral patterns they have followed in the past.[118] It also assumes that people will want the same things that other people with similar traits want. Such assumptions tie in with the notion of consumer sovereignty that Sunstein describes, but are at odds with the notion of political sovereignty.[119] Put differently, these algorithms treat individuals strictly as consumers, and not as citizens.[120] A consumer may consistently want the same type of "product," but a citizen cannot fulfill his role in a democracy if his beliefs and opinions are never challenged and hence never evolve.[121]

[36] Further, this paper argues the assumption that personalization algorithms truly represent users' choices requires a leap in reasoning. Users' attention is what drives profit, and when the information options available to each person start rising toward infinity, the best way to get a user's attention is to provide content that speaks to his "idiosyncratic

---

[117] *Id.* at 54–55.

[118] *See* Constine, *supra* note 100.

[119] *See* #REPUBLIC, *supra* note 88, at 53.

[120] *See id.* at 55 ("If we care only about consumer sovereignty, the only question is whether consumers are getting what they want – a question that seems, unfortunately, to be dominating discussions of the internet and other new technologies.").

[121] *See*, *e.g.*, John Stuart Mill, *On Liberty* at 68 (Curtis Weyant & Martin Petitt eds., 2011) ("He who knows only his side of the case, knows little of that.").

interests, desires, and needs."[122] These market forces give rise to the motto of *relevance* in Silicon Valley.[123] However, we should pause before assuming that the system is designed to honor and respect users' freedom or true choice. Showing users material they agree with is more likely to capture their attention and more likely to generate "shares" and "likes," which are a type of currency in the advertising market.[124]

[37] Successfully capturing a user's attention however is not equal to respecting that user's choice. Concepts such as bounded rationality, cognitive overload, and the idea of "least objectionable programming," can challenge such an assumption. I elaborate on these concepts in the proceeding section.

C. Users' Freedom of Choice

[38] Free choice is a notion that comes up frequently in defense of personalization algorithms. The argument suggests that if someone wants to have a narrow viewpoint and not be exposed to uncomfortable or challenging material, they have every right to do so, and the government cannot undermine the autonomy of individuals by telling them what they should or should not read and what they should or should not be exposed to. Perhaps this argument would make sense in the first version of the "Daily Me" envisioned by Negroponte, where the individual was in complete control of what went in to his personalized version of the world.[125] Even to that, critics would probably object, with arguments along the lines of Sunstein's view, pointing out the distinction between our duties as citizens and our role as a consumer, as well as the requirements

[122] Pariser, *supra* note 24, at 24.

[123] *See id*.

[124] *See* Bell et al., *supra* note 27, at 61.

[125] *See* Negroponte, *supra* note 87, at 153.

of a deliberative democracy, which ultimately guarantee our autonomy and freedom.[126]

[39] Regardless, that version of the "Daily Me" where the individual is solely responsible for designing his "experience" of the world exactly the way he wants it,[127] is at best an illusion these days. The users are clearly not in full control of the filtering.[128] My argument (and fear) is that the degree of control users will retain in a few years is minimal. In the context of social media platforms, we can visualize different levels of choice the individual can exercise. A user first chooses to join and regularly visit Facebook; then the user chooses who they become friends with and what pages they subscribe to or like. They also choose how much to disclose about themselves; their privacy settings; and finally, what to read, what to "like," and who to interact with.

[40] For simplification purposes, let's look at an isolated moment where a user signs on to Facebook to "see what is going on." Let's assume that they follow several different content sources and they have a wide range of friends. Their newsfeed however is highly filtered based on Facebook's idea of what they want to see, based on what Facebook believes is relevant to them, potentially leaving out a big part of the people and topics you have actually chosen to follow.[129] Facebook's version of what is relevant is based on data they have collected on the user, algorithmic inferences about him, their previous clicks coming from an already filtered environment, and Facebook's value judgment on what constitutes similar content or similar users.[130] When platforms claim that they are respecting

[126] *See generally* #REPUBLIC, *supra* note 88, at 54 (discussing the tension between consumer and political sovereignty).

[127] *See* Negroponte, *supra* note 87, at 153.

[128] *See* Oremus, *supra* note 86.

[129] *See id.*

[130] *See id.*

users' preferences, it is these factors that they are referring to. If we look at these factors one by one, my argument is that the notion of true choice breaks down in each of them.

[41] *First,* data collection is hardly a choice of the user. While it is true that the user chooses what to post or what photos to upload; they do not choose what others do that involves them; and they certainly do not *meaningfully* choose the tracking, inferences, and use of their data that Facebook may engage in.[131] There is no lack of support from scholars for the idea that consenting to a privacy policy has completely lost its meaning in the current information environment, and in the context of big data.[132] This is one of the biggest challenges in the field of information privacy.

[42] The main idea is that there is a false assumption of rationality in privacy decision-making, a process that is challenged by information asymmetries, externalities, and uncertainties; as well as the "bounded rationality" of humans, who in such complex situations, because of high deliberation costs and their inability to process and compute the expected utility of every alternative action, take reasoning shortcuts (*i.e*. use heuristics) that may lead to suboptimal decision making.[133] Further, consumers may find it pointless to avoid collection by one particular product or service and forgo any such effort given the vast data collection that is generally taking place.[134] Scholars also point out that the information needed to assess the "expected disutility" of any particular instance of data collection relates to *unknown* future uses or misuses of

[131] *See id.*

[132] *See* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1879, 1885 (2013).

[133] *See id.* at 1887.

[134] *See* Katherine J. Strandburg*, Free Fall: The Online Market's Consumer Preference Disconnect*, 2013 U. CHI. LEGAL F. 95, 96–97 (2013).

that information, by the data recipient or *unknown* others, which may cause *unknown* harms.[135]

[43] Further, the inferences that Facebook makes about the user may or may not be correct, and are surely not a choice of the user. The same stands in the case of Facebook's categorization as to what content is similar to content a user has previously indicated interest in, as well as what constitutes similar users that are presumed to have similar interests. More fundamentally, the notion that similar users should view similar content is certainly "efficient" from the companies' perspective, as it aligns very well with profit-maximizing goals, but using relevance as a driver of personalization also means that stimuli are tailored to play to existing inclinations and choices become narrower, most of the time without our awareness. However, one of the most defining human characteristics is that we are "*unpredictably* individual," and that is what our freedom is based on.[136] This is ignored when efficiency becomes the highest of values, leading to an appearance of free choice on the outset (in the sense of absence of coercion), but a lack of meaningful autonomy because the individual does not fully own the exploration process that brought him to the choice.

[44] Finally, clicking on something presented to me that is already based on assumptions and filtering hardly qualifies as meaningful choice for future experiences.[137] If I enter an environment of predetermined options, I may choose amongst those options, but it is very possible that I would have been more interested in something completely different, had it

---

[135] *See id.* at 130-34.

[136] *See* James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151, 1181 (2004) ("To be free was to exercise free will, and the defining characteristic of creatures with free will was that they were unpredictably individual.").

[137] *See* Sofia Grafanaki, *Autonomy Challenges in the Age of Big Data*, 27 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 803 (2017), for a comprehensive discussion on the effects of Big Data on individual autonomy and choice.

been presented to me. In a subtle process of continual feedback, every action we take reinforces the loop by continually adjusting the information environment to our comfort level, thus making it even harder to see options that would not be predictable by the algorithms, let alone chosen by them.[138] At a very fundamental level, one's capacity and facility for choice requires "a degree of freedom from monitoring, scrutiny, interference, and categorization by others."[139] This is the core of meaningful autonomy, with a broad social value as the cornerstone of a democratic society.[140] *Free choice* requires not only absence of coercion in the moment of choice, but also independence and authenticity in the process that leads to the choice.[141] In a contingent world this process inevitably involves a social context, thus the developing self has to continuously engage in boundary management between "autonomous selfhood" and the "reality of social shaping."[142]

[45] While it may sound exaggerated to think of Facebook as taking over our self-development process because a big portion of the users, at least of the ones thinking about such issues, are of a certain age where the core part of self-development has already taken place,[143] this is not necessarily the case for teenagers, who are fully immersed in technology, and interact with it as a natural part of life. A recent report from Data & Society shows that what counts as "news" for young adults is conflated

---

[138] *See* Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904, 1913 (2013).

[139] *See* Cohen, *supra* note 8, at 1426.

[140] *See id.* at 1423–27.

[141] *See* John Christman, *Autonomy in Moral and Political Philosophy*, STAN. ENCYCLOPEDIA OF PHIL.§ 1.2 (Edward N. Zalta ed., 2015), http://plato.stanford.edu/archives/spr2015/entries/autonomy-moral/, https://perma.cc/CV47-BVK3.

[142] *See* Cohen, *supra* note 138, at 1909.

[143] *See* Katheryn A. Andresen, *Marketing Through Social Networks: Business Considerations – From Brand to Privacy*, 38 WM. MITCHELL L. REV. 290, 297 (2011).

with a social element, i.e. what their friends are doing, and deeply influenced by the language and design of social networks.[144] When asked to define "news", teens and young adults identified "what's trending" and "what my friends are talking about," and when asked to share screenshots of news items, they included personal safety check-ins, and engagement announcements of friends.[145]

[46] What I hope the preceding points illustrate, is that when platforms claim they merely reflect their users' preferences (or society for that matter), they ignore the role they have in shaping the very preferences they claim to reflect. While media has always played a role in shaping society, the difference is that what used to be a clear external influence, has now become a chameleon that we hardly notice. As information on users becomes more accurate and comprehensive, and users become more "precisely and individually predictable" in terms of their reaction to external stimuli, platforms will be more efficient and invisible in their attempts to obtain particular behavioral responses.[146]

[47] Going back to the hypothetical user, proponents of personalization algorithms would argue that users do exercise choice over what they actually read, and can "correct" the algorithm by signaling out items they do or do not want to see.[147] In other words, we can get out of the loop if we *actively* want to. It would, however, require what has been called an

---

[144] *See* Laura B. Harris, *How Young Adults Define 'News': 7 good questions with Data & Society's Mary Madden*, AM. PRESS INST. (Apr. 12, 2017, 12:52 PM), https://www.americanpressinstitute.org/publications/good-questions/young-adults-news-mary-madden, https://perma.cc/23EY-VHFB.

[145] *See* Mary Madden et al., *How Youth Navigate the News Landscape*, DATA & SOCIETY 17, 23 (2017), https://kf-site-production.s3.amazonaws.com/publications/pdfs/000/000/230/original/Youth_News.pdf, https://perma.cc/E27G-SY33.

[146] Benkler, *supra* note 11, at 23.

[147] *See* #REPUBLIC, *supra* note 88, at 53–54.

"effort tax," and research has consistently shown that the "power of inertia" takes over in most cases.[148] Put differently, while we can get out of the loop if we spend the time and effort required, the practical reality is that unless we have a strong objection, we will not. This does not imply that the government should be responsible for saving us from our own inaction and complacency, or even laziness, it simply undermines the idea that consumption of content we are served equates to true preference.

[48] A quick attempt to do such "correcting" on Facebook can illustrate the point. Up to this writing, it is an extremely onerous process to curate your own feed, as it requires clicking on items one by one and either "hide post" together with "show fewer posts like this", "unfollow" the user that made the post, or see it in a tab.[149] Interestingly, it is a lot easier to set the parameters on what ads one gets, as Facebook has a part of its settings devoted to ad preferences where users can see the categorizations that apply to them and opt in or out of different categories.[150] Presumably this is a result of concerns about regulatory compliance and public perception.

[49] The point is that in an environment where we are flooded with information, we simply do not have the time to constantly question and test the filter ourselves, and for the most part, we settle with the "least objectionable alternative."[151] In fact, Pariser's filter bubble theory places significant weight on this concept.[152] He describes the theory of least objectionable programming as it originates from researching TV viewers' behavior in the 1970s, where it was noticed that with the increasing

[148] *See* CASS R. SUNSTEIN, CHOOSING NOT TO CHOOSE: UNDERSTANDING THE VALUE OF CHOICE 34, 35 (2015).

[149] *See* FACEBOOK, https://www.facebook.com/, https://perma.cc/ZJ46-TB23 (last visited Sept. 20, 2017).

[150] *See id*.

[151] *See* Pariser, *supra* note 24, at 68.

[152] *See id.*

number of available channels, "people quit channel surfing far more quickly than one might suspect."[153] "During most of those thirty-six hours a week (that Americans watch TV), the theory suggests, we're not looking for a program in particular. We're just looking to be unobjectionably entertained."[154] That said, "choosing not to choose" (*i.e.* adopting the default rule without inquiry or letting a machine make the choice for us), is a real, important, and often necessary choice that can save us from the costs and burdens of constantly making active choices.[155] Cass Sunstein's recent book provides an in-depth analysis of the value of choice, and examines whether active choices or defaults are best, as well as why and when. [156]

[50] Sunstein sees the rise of personalized default rules as a blessing that can contribute to human freedom, if used in the right way.[157] Data-driven decision making and default rules have real power and impact on our lives that we are only just beginning to understand. Notably, Sunstein also points out that we sometimes choose passively when we want to avoid the feeling of responsibility that comes with active choices, particularly when a decision has a moral dimension.[158] In those cases, where the concerns become the "risk of manipulation, compromising human agency and even dignity," he finds it imperative that the default rule is made public.[159]

---

[153] *Id*.

[154] *Id*.

[155] *See* Sunstein, *supra* note 148, at 15.

[156] *See id.* at 103.

[157] *See id.* at XIII.

[158] *See id.* at 48.

[159] *Id.* at 51.

[51] For "choosing not to choose" to be a truly autonomous choice, we need at least a minimum understanding of what it is that we are relinquishing,[160] but big data algorithms make that increasingly hard given their obscure and complex nature.

[52] On top of it all, the concept of bounded rationality of humans comes into play again when we look at the news items users decide to read. In a recent paper fueled by the "post-truth" environment of the 2016 Presidential Election, Josef Drexl poses an important question: "Do users really act rationally when they prefer false statements and conspiracy theories to verifiable facts?"[161] He suggests that users are in fact boundedly rational when they consume and share news on the Internet, stressing that "human beings suffer from confirmation bias, which leads to so-called 'motivated reasoning'",[162] a term that describes "emotion-driven reasoning that is designed to avoid emotional dissonance," and often works as a "defense to contrary evidence and often discredits the source of such evidence."[163] Put differently, the idea is that "evolution has conditioned humans to select data in a way that confirms their convictions in order to help them to win arguments."[164] It is emotionally driven because ultimately the individual *feels* better.[165]

---

[160] *See generally id.* at 135 (discussing extreme cases of total relinquishment of choice-making power in the context of "choosing not to choose").

[161] Josef Drexl, *Economic Efficiency Versus Democracy: On the Potential Role of Competition Policy in Regulating Digital Markets in Times of Post-Truth Politics*, MAX PLANCK INST. FOR INNOVATION AND COMPETITION Research Paper No. 16-16 at 10 (2016).

[162] Robert T. Carroll, *Motivated Reasoning,* THE SKEPTIC'S DICTIONARY (Oct. 17, 2015), http://skepdic.com/motivatedreasoning.html, https://perma.cc/2YRW-89BH.

[163] Drexl, *supra* note 161, at 10.

[164] *Id.*

[165] *See id.*

[53] This brings us to the final layer of choice identified, namely the choice to use the platform to begin with. One could say that if an individual is not happy with the filtered environment, they could not use Facebook and go directly to the news sources In reality, this is similar to saying that if a user of a service like Google doesn't like the privacy policies, they are always free to quit using Google. In the case of Facebook specifically, there are several accounts of users who strongly disagree with the values and direction of the company, yet still feel they need to use Facebook, because by leaving they would lose some connections which would become practically non-existent without Facebook.[166] One commentator goes as far as to analogize this effect with the typical utility company whose customers are always disappointed, yet they have no choice but to use them because of their monopoly status.[167]

[54] That said, it should be noted that there is a distinction between a user who is looking for something specific, and the user who goes on Facebook knowing that he will get the daily/hourly version of what's going on in the news and with his friends. If a user is looking for something specific they are much more likely to actively leave the filtered environment and go to news sources directly. However, most users are on Facebook just to browse and get an aggregated version of news and updates, and it is in those cases that he dangers are present.[168] In those instances, Facebook takes the role of traditional public forums, where diversity of views is of the essence.[169]

---

[166] *See* Danah Boyd, *Facebook is a Utility; Utilities Get Regulated*, APOPHENIA, http://www.zephoria.org/thoughts/archives/2010/05/15/facebook-is-a-utility-utilities-get-regulated.html, https://perma.cc/WZ24-2AN7 (last visited Sept. 18, 2017).

[167] *See id.*

[168] Darkojohnson, *What Can Psychology Tell Us About Why People Go To Facebook?* ADWEEK (Nov. 8, 2010), http://www.adweek.com/digital/what-can-psychology-tell-us-about-why-people-go-to-facebook/, https://perma.cc/MZ43-5JHH.

[169] *See generally* Sunstein, *supra* note 88.

[55] Finally, going back to the big picture I painted in Part I, the boundaries between platforms and online experiences are not clear at all. The goal of the technology companies is to provide a seamless online experience for their users, in other words one's data and behavioral patterns on one platform will most likely follow them around in some form.[170] Offline behavior will be increasingly included too. Data brokers are already aggregating information from different sources and the more intrusive technology becomes - Internet of Things and Augmented Reality devices are examples – the more data will be available to everyone.[171] Scholars call this phenomenon "context collapse."[172]The notion of these platforms as walled gardens that we can choose to enter and leave with no consequences elsewhere, is at best somewhat romantic.

D. A Call For Regulatory Intervention

[56] To summarize the argument so far, there is no doubt that filtering is necessary, as it is simply impossible to be exposed to or to consume all available information. There has always been some type of filtering going on in the media space, and the online market has now produced a particular kind of algorithmic filtering aiming to personalize the individual user experience.[173] Social media platforms such as Facebook, content distributers such as Netflix, retailers like Amazon, online media

---

[170] *See* Barry Schwartz, *New Google AdSense Auto Ads Beta*, SEARCH ENGINE ROUNDTABLE (Sep. 18, 2017), https://www.seroundtable.com/google-adsense-auto-ads-beta-24476.html, https://perma.cc/WJ5Y-KGJM.

[171] *See* Paul Boutin, *The Secretive World of Selling Data About You*, NEWSWEEK (May 30, 2016), http://www.newsweek.com/secretive-world-selling-data-about-you-464789, https://perma.cc/S42X-627D.

[172] Boyd, Danah, *How Context Collapse Was Coined: My Recollection*, APOPHENIA, http://www.zephoria.org/thoughts/archives/2013/12/08/coining-context-collapse.html, https://perma.cc/H6XE-PWKH (last visited Sept. 20, 2017).

[173] *See generally* Pariser, *supra* note 24, at 6 (discussing the changing landscape of the digital world).

aggregators, and search engines, all use such personalization algorithms; the existing versions of which have been criticized for producing "feedback loops," "filter bubbles," or "echo chambers" on the individual level, and a general "meme culture" in society.[174]

[57] To be clear, I am not arguing that we currently live in filter bubbles. At least not in their fully actualized forms that create the dystopias critics have warned about. They are present to some extent and critics disagree as to the degree.[175] We are luckily not living in the Matrix, or in Brave New World, or in the Panopticon, nor are we subject to a thought police as in George Orwell's version of 1984, or experiencing some version of Kafka's Trial. Broadcast media still play a big role. Personalization algorithms are not yet perfectly optimized so as to keep users in an inescapable filter bubble, and some parts of the world still lack Internet access, but as Bill Gates has famously said:

> *"We always overestimate the change that will occur in the next two years and underestimate the change that will occur in the next ten. Don't let yourself be lulled into inaction."*[176]
> *Bill Gates*

[58] What I am arguing is that the existing architecture of our online experiences is based on design choices driven mostly by the advertising

[174] *See* Sunstein, *supra* note 88, at 2–3; Farhad Manjoo*, How Netflix Is Deepening Our Cultural Echo Chambers*, N.Y. TIMES (Jan. 11, 2017), https://www.nytimes.com/2017/01/11/technology/how-netflix-is-depening-our-cultural-echo-chambers.html, https://perma.cc/5QLZ-39JL.

[175] *See* Kartik Hosanagar et al., *Will the Global Village Fracture Into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation*, 60 MANAGEMENT SCIENCE 805-06 (2014); *see also* Zeynep Tufekci, *How Facebook's Algorithm Suppresses Content Diversity (Modestly) and How the Newsfeed Rules Your Clicks*, MEDIUM (May 7, 2015), https://medium.com/message/how-facebook-s-algorithm-suppresses-content-diversity-modestly-how-the-newsfeed-rules-the-clicks-b5f8a4bb7bab, https://perma.cc/4VTA-PW9Z.

[176] Allison Cerra et al., TRANSFORMING BUSINESS: BIG DATA, MOBILITY, AND GLOBALIZATION 191 (2013) (quoting technology tycoon, Bill Gates).

market, and that are not aligned with the conditions required for autonomous citizens and deliberative democracies. They fit well with the idea of a user as a "sovereign consumer," but are not aligned with individuals as citizens of a democratic society. Because a significant percentage of the content that gets filtered represents political speech, algorithmic filters should be designed based on goals that are much more aligned with our notion of democracy, such as the promotion and maintenance of a functioning marketplace of ideas.[177] It is with all this background in mind, that I am calling for regulatory intervention in the social media/content aggregation space.

[59] This is not an area where we can expect the market to correct itself. As Benkler points out, what drives markets and produces "welfare" is the satisfaction of individual 'preferences' that are exogenous and preexist market relations.[178] This idea of a market however, is entirely inconsistent with the social and behavioral shaping enabled by big data methods. Benkler takes the point further and argues that while a critique of markets based on the "endogeniety of preferences" is not a new theme, the scale to which behavioral manipulation has become possible is unprecedented and has not been scientifically studied.[179]

[60] No regulation is going to perfectly fix these problems either, some of which existed before machine-learning algorithms were ever used in the media context, but at the very least, given the accelerated rate of change big data has brought, we should be feeling an unprecedented urgency to question and challenge design architectures where media filtering results

[177] *See* John Bohannon, *Is Facebook Keeping you in a Political Bubble?*, SCIENCE MAGAZINE, (May 7, 2015), http://www.sciencemag.org/news/2015/05/facebook-keeping-you-political-bubble, https://perma.cc/7JSM-FMQS (describing how Facebook utilizes certain algorithms to determine which politicized posts appear on users Facebook feeds, ultimately determining that "the jury is still out" on whether Facebook is good or bad for Democracy).

[178] *See* Benkler, *supra* note 11 at 23–24.

[179] *See id.*

in "relevance" being the highest of values. Personalization algorithms are not inherently good or bad; the design choices behind them are what drives their impact.[180] It is those design choices that I envision regulation to address, by requiring factors other than those falling under the catchphrase "relevance" to be considered, such as content being diverse, challenging, important, or serendipitous.[181]

PART II

[61] While we may not yet be fully living in a perfectly integrated perfectly tailored information system with no room to go outside algorithmic predictions, there are growing concerns that we are sleepwalking towards a future of algorithmic regulation.[182] Privacy regulations and privacy policies have not been able to prevent such algorithmic regulation, nor does it seem likely that they will, and the future is arriving faster than ever; technological developments have been showing us that in many different industries and fields.[183]

[62] As I have argued in Part I, in the media context, our previous understandings of concepts such as individual *choice* and *preference* are significantly challenged. The proliferation of content in an environment where user attention is scarce and expensive, coupled with the increased

[180] *See* Pariser, *supra* note 24, at 14–15; *see*, *e.g.*, David Talbot & Jeff Fossett, *Exploring the Role of Algorithms in Online Harmful Speech*, HARVARD KENNEDY SCHOOL, (Apr. 10, 2017) https://shorensteincenter.org/exploring-role-algorithms-online-harmful-speech/, https://perma.cc/4AK9-F7RW (discussing the ideal use of algorithms).

[181] *See* Pariser, *supra* note 24, at 12, 15.

[182] *See* Evgeny Morozov, *The Real Privacy Problem*, MIT TECH. REV. (Oct. 22, 2013), https://www.technologyreview.com/s/520426/the-real-privacy-problem/, https://perma.cc/XA3U-C83N.

[183] *See* Rita Gunther McGrath, *The Pace of Technology Adoption is Speeding Up*, HARV. BUS. REV. (Nov. 25, 2013), https://hbr.org/2013/11/the-pace-of-technology-adoption-is-speeding-up, https://perma.cc/44XB-GBJH.

integration of all our online experiences, undermines the autonomy of our choices significantly. This effect can be further accentuated if we take into account the effect surveillance capitalism has had on our online behavior.[184] Studies have shown that individuals often self-censor, both in what they read and what they write online, in an attempt to manage their digital footprints.[185]

[63] Fears of affecting future job prospects for instance, cause young adults to avoid sharing controversial content on social media.[186]

[64] Against this background, this paper proposes regulatory intervention in order to preserve First Amendment values. Broadly speaking, the type of regulation envisioned here would be designed with one key objective: to ensure and promote a *functioning* marketplace of ideas, where diverse viewpoints are available and accessible.[187] This regulatory goal is very much aligned with First Amendment values.[188] A passage from Justice Brennan's opinion in *Roth v U.S.* illustrates such values:

[184] *See* Jonathon W. Penney, *Chilling Effects: Online Surveillance and Wikipedia Use*, 31 BERKLEY TECH. L. J. 1 (2016) (discussing the results of the first empirical study providing evidence of regulatory "chilling effects" of Wikipedia users associated with online government surveillance, noting that "contrary to the 'privacy paradox,' privacy concerns are being reflected in online behavior").

[185] *See generally* Manya Sleeper et. al., *The Post that Wasn't: Exploring Self-Censorship on Facebook*, CARNEGIE MELLON UNIV. (2013) http://dl.acm.org/citation.cfm?id=2441865, https://perma.cc/X5AE-KUH5 (explaining a study describing social networking user's decisions to share or not share content on their media profiles).

[186] *See* Madden, *supra* note 145.

[187] *See* Mill, *supra* note 121.

[188] *See* Roth v. United States, 354 U.S. 476, 484–88 (1957).

"The protection given speech and press was fashioned to assure unfettered interchange of ideas for the bringing about of political and social changes desired by the people."[189]

[65] To clarify, I am not implying that a functioning marketplace of ideas means that the truth necessarily prevails. That is a separate issue and regardless of its value as an ideal, the courts have made clear that the dominance of truth is by no means a requirement for free speech.[190] At the same time, the assumption that simply protecting against censorship is tantamount to a guarantee of free expression is contestable, and technological changes challenge the assumption even further.

[66] In some ways, we have seen all this before. Writing in 1967, Jerome Barron in his classic article *Access to the Press*, rejected the "romantic" view of free expression, and argued for a right of access that would ensure that ideas can enter the marketplace and be accessible.[191] Concerns about access to the press may sound irrelevant in our current information environment where citizens can easily self-publish their views and ideas, but there is romanticism and naïveté here too if we assume that because information is somewhere online and can theoretically be accessed by anyone, this naturally results in an uninhibited marketplace of ideas. Rewriting in 2008, Barron acknowledges that his originally envisioned right of access was conceived at a very different time, yet concludes once more: "[p]aradoxically, the belief that a contemporary marketplace of ideas exists has become the rationale from preventing it

---

[189] *Id*. at 484.

[190] *See*, *e.g.*, American Booksellers v Hudnut, 771 F.2d 323, 330-31 (7th Cir.1985) ("A power to limit speech on the ground that truth has not yet prevailed and is not likely to prevail implies the power to declare truth…If the government may declare the truth, why wait for the failure of speech? Under the First Amendment however there is no such thing as a false idea …, so the government may not restrict speech on the ground that in a free exchange truth is not yet dominant.").

[191] *See* Jerome A. Barron, *Access to the Press—A New First Amendment Right*, 80 HARV. L. REV. 1641, 1668 (1967).

from becoming a reality."[192] The point here is that these are not new problems and the law has addressed them before; the difference is in the players involved and their respective interest and power.[193]

[67] Having as an objective the "unfettered interchange of ideas," the envisioned regulation would have to do more than require disclosures, disclaimers, or explanatory notices.[194] This is because the level of behavioral persuasion (if not manipulation) that Big Data has enabled would undermine any effect disclosures could have, much like what happens in the context of privacy notices.[195] Moreover, requiring transparency has inherent problems when we are dealing with machine learning algorithms, as it is unclear what types of explanations would satisfy a "transparency" requirement, and whether they would actually serve the desired purpose. Regulation would therefore, at least to some degree, interfere with the filtering algorithms, in that it would target the criteria on which the algorithms base their filtering process, the most obvious being the notion of "relevance" and its definition, as well as the weight of the particular definition on the overall information filtering.

[68] Regulation in the media space is not a novel proposal, in fact scholars such as Sunstein, argue that avoiding regulation in the communications market is completely unrealistic and the only question is what kind of regulation we should have.[196] The type of regulation becomes especially important when assessing whether regulatory intervention

---

[192] Jerome A. Barron, *Access Reconsidered*, 76 GEO. WASH. L. REV. 826, at 844 (2008).

[193] *See generally,* Rebecca Tushnet , *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986, 986-87 (2008) (discussing how Barron's argument remains relevant in a changing communications landscape).

[194] *See* Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms,* 55 B.C. L. REV. 93, 94, 106 (2014).

[195] *See id.*

[196] *See* Sunstein, *supra* note 88, at 178.

would pass First Amendment scrutiny. While precise details of specific regulatory proposals will not be addressed, it is their nature that is important in assessing their constitutionality.

[69] The remainder of this paper proceeds to identify and evaluate potential objections to such regulatory intervention. I view potential objections as challenging the constitutional basis for introducing regulation, based on the First Amendment rights of the private entities - platforms- that use the algorithms. Such *free speech* type objections would suggest that regulation would intervene with protected speech in a way that is unconstitutional.

A. Free Speech: Platforms and their First Amendment Rights

[70] Free Speech objections to regulatory intervention concern the private entities that use the algorithms, i.e. the platforms, and come on two levels: the actual content (such as particular news items) as speech that should not be regulated, and the machine-learning algorithm as speech that should not be regulated.[197] In other words, there are two separate layers of free "speech" that could be abridged: the actual (news or other) items/posts presented to users as "speech," and each platform's machine learning personalization algorithm as "speech."[198]Proponents of characterizing the personalization algorithm as protected speech argue that it deserves First Amendment protection because it expresses and represents the respective platform's point of view about what is most "meaningful" or "relevant" to its users.[199]

[71] As briefly described above, the envisioned regulation would interfere with what machine-learning personalization algorithms of the

[197] The idea that the machine learning algorithm is speech for First Amendment purposes is itself much debated, *see infra* discussion and notes in Part II - ii.

[198] *See id*.

[199] *See id.*

platforms are doing, and would have an effect on the content that platforms present to users. Viewed with a First Amendment doctrine lens, this type of regulation would be categorized as content-based, but viewpoint neutral,[200] because regulation would be directed to the content of speech, but would not seek to ban or promote a particular viewpoint. When it comes to the machine-learning algorithm as speech (if it can be characterized as speech), however, regulation requiring users be exposed to diverse opinions would be deemed as viewpoint discrimination.[201]

[72] Before diving into the details of the arguments, it is useful to take a step back and note that the First Amendment is not absolute.[202] The government is allowed to, and does, regulate several types of speech, such as defamatory speech or computer viruses in the software category.[203] Reasons for doing so vary depending on the category of unlawful speech, but for the purposes of the argument put forth in this paper, the "exception" of copyright law as whole can be illustrative.[204] One way to view copyright law as compatible with the First Amendment is to argue that it has built-in safeguards such as the idea/expression distinction, and the fair use doctrine.[205]

[73] For instance, since copyright does not protect ideas, but rather the expression of ideas, one can argue that the marketplace of "ideas" is not really affected by copyright laws. That said, the Supreme Court clearly

---

[200] *See id*.

[201] *See id.*

[202] *See* Sunstein, *supra* note 88, at 192.

[203] *See id.*

[204] *See id.* at 195.

[205] *See* Lee Ann Lockridge, *The Myth of Copyright's Fair Use Doctrine as a Protector of Free Speech*, 24 SANTA CLARA COMPUTER & HIGH TECH. L.J. 31, 36–7 (2007).

views copyright as "the engine of free expression,"[206] and protection of free expression is at the heart of the First Amendment.[207] In the words of Justice O'Connor in *Harper & Row v. Nation Enterprises*, "it should not be forgotten that the Framers intended copyright itself to be the engine of free expression. By establishing a marketable right to the use of one's expression, copyright supplies the economic incentive to create and disseminate ideas."[208]

[74] While the above description is a clear oversimplification of a much larger debate, copyright law is a clear illustration of the notion that the First Amendment is not an absolute, because unlike cases where physical harm is in question, the reasons for regulating speech in the area of copyright are much closer to the ideas put forward by this paper, namely promoting and protecting a vivid and functioning marketplace of ideas.

[75] The subsequent sections address each layer of speech separately. They examine similar types of regulation that have been held as constitutional in the past, and use them as support of the constitutionality of the envisioned type of regulation in the context of news aggregator and social media platforms.

i. First Amendment concerns on Content Regulation

[76] In regards to the first category of platform speech, involving the actual content items presented to users, it can be illustrative to examine the controversial, and now abolished Fairness Doctrine, a Federal Communications Commission (FCC) promulgated rule that came into effect in 1949.[209] The doctrine had two prongs, requiring radio and

[206] Harper & Row, Publrs. v. Nation Enters., 471 U.S. 539, 558 (1984).

[207] *See id.* at 559 (citing Estate of Hemingway v. Random House, Inc., 23 N.Y. 2d 341, 348 (1968)).

[208] *See id.* at 558.

[209] *See* Red Lion Broad. Co. v. Fed. Commc'ns Comm., 395 U.S. 367, 377 (1969).

television licensees to: (1) cover important controversial issues of public concern, and (2) provide a reasonable opportunity for a balanced presentation of contrasting viewpoints on such issues.[210] The Fairness Doctrine was challenged on First Amendment grounds, but upheld as constitutional by the Supreme Court in *Red Lion Broadcasting Co. v. FCC*, based on the notion that the purpose of the First Amendment was to "preserve an uninhibited marketplace of ideas," and that purpose was furthered by the Doctrine.[211] The Fairness Doctrine was later abolished by the FCC in 1987.[212]

[77] Before looking at the Doctrine more closely, it is important to clarify the objective in doing so: the purpose of the analysis is not to suggest that the Fairness Doctrine should be resurrected. It is mostly irrelevant and inapplicable in the current information landscape, and its effect is questionable. One of the reasons the FCC abolished the Fairness Doctrine was because instead of producing more speech, it actually had "chilling effects" on speech[213] because broadcasters did not want to find themselves awarding unpaid time to dissenting/contrasting viewpoints, as was required by the second prong of the doctrine, and chose instead to

---

[210] *See* Barron, ACCESS RECONSIDERED, *supra* note 192, at 827 n.7 (2008) (citing Editorializing by Broad. Licensees, 13 F.C.C. 1258); *Federal Communications Commission*, ENCYCLOPEDIA.COM (Sept. 18, 2011, 9:11 AM), http://www.encyclopedia.com/arts/news-wires-white-papers-and-books/federal-communications-commission, https://perma.cc/Z5Q2-632C (citing Excerpt of "The Fairness Doctrine").

[211] Red Lion, 395 U.S. at 375, 390, 392. For further discussion, *see* Barron, ACCESS RECONSIDERED, *supra* note 191, at 826-829.

[212] *See* Robert Hershey Jr., *F.C.C. Votes Down Fairness Doctrine in a 4-0 Decision*, N.Y. TIMES (Aug. 5, 1987), http://www.nytimes.com/1987/08/05/arts/fcc-votes-down-fairness-doctrine-in-a-4-0-decision.html, https://perma.cc/67NS-UB5L (last visited Sept. 12, 2017).

[213] *See* Thomas W. Hazlett & David W. Sosa, *Was the Fairness Doctrine a "Chilling Effect"? Evidence from the Postderegulation Radio Market,* 26 J. OF LEGAL STUD. 279, 280 (1997).

present less controversial issues.[214] For others, like Jerome Barron, the value of the doctrine was not in its enforcement (which he saw as hardly rigorous), but in its mere existence as a cautionary reminder against "excessive one-sidedness in the presentation of public affairs," that forced a sense of responsibility on broadcasters.[215]

[78] Put differently, "[t]he system didn't always work, but it did set some limits."[216]

[79] Regardless of the Fairness Doctrine's actual effect, and whether it was a good or bad policy,[217] it is the idea behind it, as a justification for viewpoint neutral, content-based speech regulation, that matters for our purposes. The idea was to ensure citizens are exposed to different and contrasting viewpoints,[218] and it was therefore, a viewpoint neutral regulation for First Amendment purposes, which was upheld by the Supreme Court as constitutional.[219] Some scholars view the *Red Lion* decision as support for the idea that the First Amendment has an affirmative dimension, i.e. that "law could not only protect freedom of expression but facilitate it."[220] Whether the broader idea holds or not, the argument here is much narrower. The claim is that if the circumstances

[214] *See id*.

[215] Jerome A. Barron, *Access to the Media: A Contemporary Appraisal*, 35 HOFSTRA L. REV. 937, 943 (2007) [hereinafter Barron, ACCESS TO THE MEDIA].

[216] Paul Krugman, *In Media Res,* N.Y. TIMES (Nov. 29, 2002), http://www.nytimes.com/2002/11/29/opinion/in-media-res.html, https://perma.cc/VA8Y-LJZL (last visited Sept. 28, 2017).

[217] *See* Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 19 (2004).

[218] *See id.* at 5.

[219] *See* Red Lion Broad. Co v. Fed. Commc'ns Comm., 395 U.S. 367, 375 (1969).

[220] Barron, ACCESS TO THE MEDIA, *supra* note 215, at 938.

and government rationales are analogous, *Red Lion* implies that the First Amendment would not be a bar to regulation of social media platforms.[221]

[80] Arguing for an analogy as to rationale and circumstances that justify speech regulation, I see two themes as central: an underlying commitment to deliberative democracy (rationale), and the notion of scarcity (circumstances).

a. Democratic Commitment as a Rationale for Free Speech

[81] The democratic roots of the free speech principle are central to the analysis. Sunstein's work is illustrative here too, as he points out a fundamental distinction between justifying free speech on a notion of consumer sovereignty, as opposed to a commitment to democratic deliberation.[222] The First Amendment values I have been referring to relate to this commitment. On that basis, the notion that political speech is more valuable than commercial speech, and therefore deserves stronger protection, is almost intuitive. Sunstein further explains that too much focus on government censorship can produce blind spots.[223] A well-functioning democracy and system of free expression have, as necessary preconditions, both an element of serendipity in the materials one is exposed to (what Sunstein calls chance encounters), as well as shared experiences for citizens, which provide a form of social glue that holds a society together.[224]

---

[221] *See* Rebecca Tushnet*, Power Without Responsibility: Intermediaries and the First Amendment,* 76 GEO. WASH. L. REV. 101, 115 (2008) (citing Jerome A. Barron, *Structural Regulation of the Media and the Diversity Rationale*, 52 FED. COMM. L.J. 555 (2000).

[222] *See* Sunstein, *supra* note 88, at 195–96.

[223] *See id*. at 193.

[224] *See id.* at 7.

b. Scarcity

[82] A main factor behind the *Red Lion* Court's reasoning was the scarcity rationale.[225]

[83] The idea was that broadcast spectrum was a scarce commodity, in that there were very few frequencies available for radio and television stations.[226] If spectrum was scarce, then there were only limited opportunities for diverse viewpoints to be expressed, which in turn justified government intervention in the marketplace of ideas.[227] This type of scarcity rationale can at first seem largely irrelevant in the current information landscape, as there is hardly scarcity of available sources. We are living in a time where technological developments have made available information practically abundant, both in terms of access to existing information, given the vast number of platforms and content sources, and in terms of creation of new information.[228] The rise of the so-called "citizens journalism" is a testament to the latter effect.[229]

[84] In the current information environment, however, there is another scarce commodity: our attention.[230] That is what is being sold to the highest bidder. As early as 1971, Herbert Simon observed that:

---

[225] *See Red Lion Broad. Co.,* 395 U.S. at 400–01.

[226] *See id*. at 401 n.28 (noting that the rationale is based on scarcity in economic sense, not technological (i.e. the FCC could open more frequencies but that was irrelevant)).

[227] *See* JOHN W. BERRESFORD, THE SCARCITY RATIONALE FOR REGULATING TRADITIONAL BROADCASTING: AN IDEA WHOSE TIME HAS PASSED, FCC Staff Research Paper, at 3 (March 2005).

[228] *See id.* at 12–13.

[229] *See* Kate Bulkley, *The Rise of Citizen Journalism*, THE GUARDIAN (June 10, 2012), https://www.theguardian.com/media/2012/jun/11/rise-of-citizen-journalism, https://perma.cc/CWD3-2G4C (last visited Sept. 20, 2017).

[230] *See* Timothy Taylor, *Economics of Information Overload*, CONVERSABLE ECONOMIST (Aug. 17, 2015), http://conversableeconomist.blogspot.com/2015/08/economics-of-

> "in an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."[231]

[85] Tim Wu picks up on this and explains how this phenomenon plays out in today's information environment in his latest book, *The Attention Merchants: The Epic Scramble to get inside our heads.*[232] The title alone gives a sense of just how valuable of a commodity our attention has become. Psychologists and neuroscientists now talk of "cognitive overload", a phenomenon where our brains get overwhelmed by too much information and too many choices, and effectively freeze, leading to indecisiveness, bad decisions, and stress.[233]

[86] Viewed this way, the abundance of content sources and platforms is almost irrelevant, because we have a limited attention span, and even without scarcity of platforms and news aggregators, if all of them (albeit independently from one another) filter content based on what is ***relevant*** to us, our perspective becomes quite narrow.[234] It is not a perfect analogy

information-overload.html, https://perma.cc/JT6Y-6ETZ (quoting Herbert A. Simon, *Designing Organizations for an Information-Rich World* in Greenberger, Martin, *Computers, Communication, and the Public Interest*, THE JOHNS HOPKINS PRESS, 40–41 (1971)).

[231] *Id*.

[232] *See* TIM WU, THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS 6 (1st ed. 2016).

[233] Margarita Tartakovsky, *Overcoming Information Overload*, PSYCHCENTRAL (January 21, 2013) https://psychcentral.com/blog/archives/2013/01/21/overcoming-information-overload/, https://perma.cc/PTZ5-ZQYT (last visited Sept. 20, 2017).

[234] *See id*.

because the scarcity now is on the demand side (the audience), whereas before it was on the supply side (spectrum/broadcast channels).

[87] This comes as a result of technical structures that shifted the power dynamics, but does not change the effect on the marketplace of ideas reasoning.[235] When the *Red Lion* Court upheld the Fairness Doctrine, it stressed it is the right of the viewing and listening public, and not the right of the broadcasters, which is paramount.[236] A functioning marketplace of ideas requires diverse views to be heard;[237] if they are not heard, regardless of whether it is due to limited access in the way that Barron first talked about,[238] or to limited possibilities for attention that make it impossible to become aware of all (or at least diverse) viewpoints, the result is the same While a strict reading of the First Amendment does not directly protect the marketplace of ideas as such, its basis is the ideal of a functioning marketplace of ideas. Justice Brennan in *NY Times v. Sullivan* made this exact point when he said that the whole purpose of the First Amendment was to encourage uninhibited and robust debate.[239] The Court in *Red Lion* rejected a reading of the First Amendment that would result in inhibiting rather than protecting this ideal, and it is on that basis that I view the analogy as valid.[240]

But are they speakers or intermediaries?

---

[235] *See* Manjoo, *supra* note 4.

[236] *See Red Lion Broad. Co.,* 395 U.S. at 390 (1969).

[237] *See* Mill, *supra* note 121, at 37.

[238] *See* Barron, *Access to the Press*, *supra* note 191, at 1641–42.

[239] *See* New York Times Co. v. Sullivan, 376 U.S. 254, 270, 279 (1964).

[240] *See Red Lion Broad. Co*., 395 U.S. at 388–89 (1969).

[88] So far, the discussion has been based on the assumption that the platforms and aggregators are in fact endorsing their content as their own speech. In reality, the content on such platforms does not represent the platforms' views, ideas, or opinions at all; to the contrary, social media platforms like Facebook consistently deny being media companies, and make sure to stay on the side of intermediaries or mere conduits.[241] This enables them to claim the protection of section 230 of the Communications Decency Act, which, under certain conditions, shields them from potential liability stemming from user-generated content.[242] That said, these types of intermediaries have been intensely criticized as suffering from a convenient "identity crisis," in that depending on what is more advantageous to them in a given situation, they will choose to be speakers when they want to protect their algorithms, but will claim they are intermediaries when they want to be shielded from liability for unlawful third-party or user generated content.[243] In the words of R. Tushnet, "[c]urrent law often allows Internet intermediaries to have their free speech and everyone else's too."[244] It is on that double identity that the preceding discussion is based, meaning that because social media platforms switch sides between being speakers or being technology companies acting as conduits, depending on what their economic interests dictate, it is useful to address both "identities," at least until courts or the legislature clarify the correct answer. It is unlikely that companies like

[241] *See* Jeff John Roberts, *Why Facebook Won't Admit it's a Media Company*, FORTUNE (Nov. 14, 2016), http://fortune.com/2016/11/14/facebook-zuckerberg-media/, https://perma.cc/NC8A-G5NB.

[242] *See* 47 U.S.C.S. § 230(c) (LexisNexis 2017).

[243] *See* Frank A. Pasquale, *Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power,*17 THEORETICAL INQUIRIES IN L. 487 (2016); *See also* Ardia, David S., *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, 377-80 (2010); Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986, 986-88, 1002, 1008-09 (2008).

[244] *See* Tushnet, *supra* note 243, at 1002.

Facebook or Snapchat will choose to be categorized as speakers or editors, as that would open them up not just to liability for unlawful speech, but also potentially subject them to other types of regulation, such as competition laws as they apply to companies in the communications space.[245] However, it is at least plausible, that if their economic interests dictate, some intermediaries may choose to be speakers with the added burden of engaging in more robust content "policing" for unlawful content, in order to preserve their profit making algorithms. In other words, it is at least conceivable, that in some cases it may be cheaper to build the tools (or employ the necessary personnel) to monitor for unlawful content, or to even pay the price of liability, than to sacrifice the advertising dollars stemming from the profit maximizing algorithms. It is also possible that the objections to regulation are presented as coming from speakers themselves (meaning the originators of content), rather than intermediaries that may or may not be speakers.

[89] First Amendment objections are of course much stronger when coming from the content producers, or if intermediaries take on an editing role beyond the safe harbor of section 230.[246] In a much quoted passage from *Columbia Broadcasting Systems v. Democratic National Committee* where a First Amendment right of access was denied, Chief Justice Burger stated, "editing is what editors are for; and editing is selection and choice of material."[247] The Court did however note that a right of access to broadcast journalism was not necessary because the Fairness Doctrine was available.[248] However, when it comes to print media, the sense that government intervention with editorial judgment is inconsistent with the

[245] *See* Pasquale, *supra* note 243, at 503.

[246] *See* Methaya Sirichit, *Censorship by Intermediary and Moral Rights: Strengthening Authors' Control Over Online Expressions Through the Right of Respect and Integrity*, 1 J. L., TECH., AND PUB. POL'Y, 54 (2015).

[247] Columbia Broad. Sys., Inc. v. Democratic Nat'l Comm., 412 U.S. 94, 124 (1972).

[248] *See id.* at 111.

First Amendment is even stronger, and issues of concentration of media owners have not convinced the courts to the contrary.[249]

[90] Such precedents may at first appear to weaken an argument for regulation. On a closer look however, I view them as potential support. In *Columbia Broadcasting*, access was denied exactly because the Fairness Doctrine was available.[250] This could imply that such a right would in fact be necessary in the absence of the Doctrine. At a minimum, it can be seen as a validation of what the Doctrine was protecting. In the case of print media and the *Tornillo* case, there was an assumption that newspapers are not subject to the scarcity problem that broadcasters are, so the public could, at least in theory, have access to diverse ideas in the marketplace, even if they did not all come from the same newspaper.[251] This may sound even more true today given the almost infinite sources of information in our networked world, but the scarcity of the public's attention, coupled with the integration of our "online" experiences, challenge this view.

[91] In other words, while yes, it is true that I can leave Facebook's filtered environment and go to another (aggregator) website, that website most likely also filters content based on relevance. In practice, visiting a different website does not look at all similar to reading a different print newspaper back in the day. Different editors back then meant (at least in theory) different values, different ideas, and some element of diversity. In today's online environment, different "editors" are different machine learning algorithms with somewhat similar values (*relevance* is key) that have nothing to do with the values we traditionally associate with editors.[252] Their main difference is their position in the constant race to

[249] *See* Miami Herald Publ'g Co. v. Tornillo, 418 U.S. 241, 252 (1974).

[250] *See Columbia Broad. Sys., Inc*., 412 U.S. at 111.

[251] *See Tornillo*, 418 U.S. at 252.

[252] *See* Org. for Econ. Co-operation and Dev. [OECD], *Algorithms and Collusion – Background Note on the Secretariat*, at 1, 13, DAF/COMP(2017)4 (June, 9 2017), https://one.oecd.org/document/DAF/COMP(2017)4/en/pdf, https://perma.cc/U6GR-CLBQ.

acquire more data than their competitors, and use that data to become even more relevant to their users.[253]

[92] Back then, there was also a notion that print was a much less intrusive medium than broadcast mediums were, but it's doubtful we could see Facebook as analogous to the less intrusive print medium. If anything, such platforms are much more intrusive than TV ever was. The level of intrusiveness of the medium determines its impact on society and the marketplace of ideas.[254] As Marshall McLuhan has famously observed, "the medium is the message," meaning that each medium stimulates different degrees of engagement and participation.[255] He was writing about television at the time and arguing that images are almost mesmerizing us to the point where their actual content becomes irrelevant.[256]

ii. First Amendment concerns for regulation of machine-learning algorithms

[93] Starting from the premise that "editing" is protected speech and "what editors are for" as discussed above, we need to assess whether the machine-learning algorithm is an "editor" in the First Amendment sense.[257] While it may seem counter-intuitive, the idea that computer code is a type of speech has received considerable support.[258] As a starting point, let's note that the reason computer software can recieve copyright protection, is that it is classified as a "literary work," traditionally

[253] *Id.* at 6.

[254] *See Columbia Broad. Sys., Inc.*, 412 U.S. at 183–84.

[255] *See* MARSHALL MCLUHAN, UNDERSTANDING MEDIA 1 (1964).

[256] *See id.* at 2.

[257] *See Columbia Broad. Sys. Inc.,* 412 U.S. at 124–25.

[258] *See* Eugene Volokh & Donald M. Falk, *GOOGLE First Amendment Protection for Search Engine Search Results,* 8 J.L. ECON. & POL'Y 883, 886 (2012).

protected by copyright.[259] The notion that the First Amendment protects algorithms was first made in the context of Google's search algorithm, and two federal court decisions have held that "search results, including the choices of what to include in those results," receive full First Amendment protection.[260] The idea is that just as the First Amendment protects newspaper editors who cannot be compelled to publish or not publish a particular content item, the same applies to search engines, who cannot be compelled to include or exclude certain links.[261] The underlying assumption is that it is the programmers who wrote the code receive the constitutional protection as speakers, and the algorithm represents their speech, which is protected. This notion has also been strongly contested; opponents warn against the possibility that the trillions of invisible decisions made by machines on a daily basis can be protected speech.[262] This could put technology companies in a position to argue against privacy or antitrust regulation using a constitutional pretense. Taking a somewhat intermediate position, others views the debate as a balancing exercise between the rights of platform owners, platform users, and the public interest, reminding us that the First Amendment "does not disable the government from taking steps to ensure that private interests not restrict, through physical control of a critical pathway of communication, the free flow of information and ideas."[263]

---

[259] United States Copyright Office*, Copyright Registration for Computer Programs*, CIRCULAR 61 (2012) https://www.copyright.gov/circs/circs61.pdf, https://perma.cc/DYX5-8RJN.

[260] Volokh, *supra* note 258, at 886. *See also* Search King v. Google Tech., Inc., 2003 U.S. Dist. LEXIS 27193 1, *11–12 (W.D. Okla. 2003).

[261] *See supra* Volokh note 258, at 886. *See also*, Langdon v Google, 474 F. Supp. 2d 622, 630 (U.S. Dist. Del. 2007).

[262] *See* Tim Wu, *Free Speech for Computers?*, N.Y. TIMES (June 19, 2012), http://www.nytimes.com/2012/06/20/opinion/free-speech-for-computers.html, https://perma.cc/7PKW-GYJ7.

[263] Frank Pasquale, *Automated Arrangement of information: Speech, Conduct, and Power*, CONCURRING OPINIONS (June 25, 2012), https://concurringopinions.com/archives/2012/06/automated-arrangement-of-

[94] Regardless of the correct answer, social media companies and news aggregator platforms have not yet made an argument that their filtering algorithms represent protected speech. Presumably this is because it would be a double-edged sword. If their algorithms are speech then they are seen as editors, just like the Google search algorithm in *Langdon*; but if they are editors, then they are media companies, and they really do not want to be media companies.[264]

[95] These social media companies and new aggregator platforms want to be seen as technology companies. Their predominant arguments for not being media companies point to the fact that they do not produce any original content but merely distribute it, and to the notion that there is no human judgment present in the filtering that would resemble the traditional editor, but rather algorithms that simply *reflect* what users want.[265] Commentators have however, been quick to point out that content production was never a necessary requirement for a media company, and *distribution* alone was always a defining characteristic.[266] As to the notion that the technical personnel that write the code do not have similarities with traditional editors, it was recently revealed that Facebook hired journalists to train their algorithms for their trending news section, but interestingly, they were not hired as employees but as independent contractors.[267] This arguably shows a caution to avoid any grounds for a

information-speech-conduct-and-power.html, https://perma.cc/RW2P-FTKN, (quoting Turner Broad. Sys. v. Fed. Commc'ns Comm., 512 U.S. 622 (1994)).

[264] *See Search King, Inc.*, 2003 U.S. Dist. LEXIS 27193 at *11–12.

[265] *See* Philip M. Napoli & Robyn Caplan, *When Media Companies Insist They're Not Media Companies and Why It Matters for Communications Policy* 8, 12 (Presentation Material for the 2016 Telecommunications Policy Research Conference in Arlington Virginia on March 18, 2016). https://ssrn.com/abstract=2750148, https://perma.cc/6YYZ-RUYM.

[266] *See id.* at 9.

[267] *See* Michael Nunez, *Want to Know What Facebook Really Thinks of Journalists? Here's What Happened When It Hired Some*, GIZMODO (May 3, 2016),

characterization other than a technology company that acts as an online intermediary.

[96] What I hope the preceding discussion illustrates is that any objection to regulation of filtering algorithms on First Amendment grounds will inevitably strengthen the case for such platforms being media companies. If they are found to be media companies, they can inevitably be subject to the public interest regulations governing media companies.[268] Regardless, if we do accept that filtering algorithm is speech, we need to evaluate First Amendment concerns against regulating it.

[97] The arguments are very similar to the discussion in section i relating to the actual content. The platforms however could try to argue that there is viewpoint discrimination when directing the regulation to the actual algorithm. The idea would be that if, for example, the Facebook algorithm represents a point of view that user A prefers to see content type X, then requiring Facebook to show user A content A-Z or a somewhat diverse content, represents a different point of view and directly targets Facebook's original belief. However, even if we accept that as speech, it is surely not the same type of editorial speech the courts were protecting in *Columbia Broadcasting* and *Tornillo.*[269] In those cases editorial speech represented the editor's judgment about what were issues of public importance, as well as his way of adhering to journalistic ethics,[270] which was classified as political speech, and the First Amendment protects it strongly.[271]

---

http://gizmodo.com/want-to-know-what-facebook-really-thinks-of-journalists-1773916117, https://perma.cc/77TU-K2BM.

[268] Napoli & Caplan, *supra* note 265, at 9.

[269] *See generally* Miami Herald Pub. Co. v. Tornillo 418 U.S 241 (1974); Columbia Broadcasting System, Inc. v. Democratic Nat'l Comm., 412 U.S 94 (1973) (holding that editorial acts in journalism and broadcasting fall under First Amendment protection of free speech).

[270] *See Tornillo*, 418 U.S. at 258; *Columbia Broad. Sys., Inc.* 412 U.S. at 111.

[98] In the case of Facebook, even if the actual content selected is political in nature, the algorithm itself is much closer to commercial speech. It is not even directed to more than one person. It sounds a lot more like a salesperson in a department store observing a customer, *noting* his/her gender, accent, clothes, shoes, watch, handbag, jewelry, and general style, *inferring* age and social and economic background, and then *directing* that customer to items the salesperson thinks they would be interested in, based on their experience or knowledge of what other customers with similar (observed or inferred) traits were interested in. Whether the customer is buying shampoo or a book should not make much difference if the salesperson method is the same. He/she is "selling," not "editing," thus there is no particular viewpoint or message that they have so as to claim First Amendment protection. Instead, their "speech" is driven entirely by the goal of completing a sale. Under First Amendment doctrine, that is "speech 'which does no more than propose a commercial transaction'," and thus deserves a lower degree of protection under the First Amendment.[272] Advertising that simply "links a product to a current public debate" is not entitled to the same constitutional protection afforded to noncommercial speech.[273] If the speech is not misleading, it must be determined whether the governmental interest to be served by the restriction on commercial speech is substantial, whether the regulation directly advances the governmental interest asserted, and whether it is not more extensive than necessary to serve that interest.[274]

[99] If we tried to apply this test to personalization algorithms, the issue of whether they are misleading is at least debatable, but hard to examine

[271] *See Tornillo*, 418 U.S. at 258.

[272] Virginia State Bd. of Pharmacy v. Virginia Citizens Consumer Council, 425 U.S. 748, 762 (1976), (quoting Pittsburgh Press Co. v. Human Relations Comm'n, 413 U.S. 376, 385 (1973)).

[273] Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 563 (1980).

[274] *See id.* at 564.

since we cannot identify their exact message. It is not even stable enough to describe, as the algorithm continuously updates with every new input/interaction of the user.[275] Surely one could argue they are misleading if they are directing users to fake news, but are they misleading by claiming that they are showing users content that is relevant to them? Arguments here could be valid on both sides. Regardless, the promotion and maintenance of a functioning marketplace of ideas would surely qualify as a substantial government interest, leaving the question of whether the restriction is more extensive than necessary to serve that interest. My claim is that even without describing the regulation in detail, I have illustrated that some kind of interference with the algorithm is necessary. What I have proposed in abstract is to include values that are aligned with democratic principles and a functioning marketplace of ideas. I suggested values such as diversity and content that is important, different, challenging, or serendipitous. Ultimately it depends on how these ideas are implemented, a task that would require engineers and legislators to work together, but it is not a stretch to envision that the goal can be achieved without making the regulation "more extensive than necessary" to serve the governmental interest.[276]

[100] From a different perspective, one could arguably say that the algorithm's viewpoint is not just about one person at a time, but rather represents Facebook's belief that its users in general prefer to see content that is relevant to them. However, again, this would only mean that the salesperson in my previous example was acting under a general mandate from the department store, and it wouldn't change the nature of the speech. That is not to say that the salesperson does not have the right to say whatever he/she wants to the customer. What I am trying to illustrate is that this is not the highest level of speech protected by the First Amendment (political speech),[277] but rather it is viewed as lower value

[275] *See* Oremus, *supra* note 86.

[276] *Cent. Hudson Gas*, 447 U.S. at 591.

[277] *See* Sunstein, *supra* note 88, at 205.

commercial speech.[278] The First Amendment concerns and objections to regulation are therefore much weaker.

What if they are technology companies?

[101] Turning to the platforms' other identity, we now look at the implications if they are not speakers. If all they are is providers of an "interactive computer service" to use the language of section 230, then the First Amendment objections do not make much sense.[279] So could the FCC or Congress regulate the way they distribute content?

[102] The FCC has for years stayed away from Internet regulation, but there has been a recent shift in its approach, indicating that it has begun to consider the Internet as falling under its broader regulatory authority for communications services.[280] This shift is evidenced by its intervention on the net neutrality issue .[281] The concept of net neutrality was born out of an idea of the Web as a public good or public place, which would require fair access if it were to continue to exist as such.[282] That lead to the FCC reclassifying broadband providers as telecommunications services, giving itself the authority to regulate them as public utilities.[283] Interestingly, scholars and commentators have begun to argue that Facebook and social media platforms in general are a utility.[284]

---

[278] *Id.*; *see also,* STEVEN G. BRODY & BRUCE E. H. JOHNSON, ADVERTISING AND COMMERCIAL SPEECH: A FIRST AMENDMENT GUIDE § 2.1 (2d ed. 2016).

[279] 47 U.S.C. § 230 (2012).

[280] *See* Napoli & Caplan, *supra* note 265, at 4–5.

[281] *See id.* at 5.

[282] *See id.* at 18–19.

[283] *See id*. at 5.

[284] *See* Mark Andrejevic, *Public Service Media Utilities: Rethinking Search Engines and Social Networking as Public Goods*, 146 MEDIA INT'L AUSL. 123, 129 (2013); Dana Boyd, *Facebook is a Utility; Utilities Get Regulated*, APOPHENIA (May 15, 2012),

[103] Regardless of the validity of that argument, there is certainly support for the notion that information is a public good, leading commentators to argue for notions of fairness in its distribution.[285] Introna and Nissenbaum for instance, view the Web as a "special kind of public place," that fulfills more than the functions of traditional public spaces (like museums, parks, beaches and schools), such as serving as a medium for artistic expression, space for recreation, a place for exhibiting items of historical and cultural importance, and being able to educate. [286] Most importantly, it functions as a conveyor of information, with access to such being construed as a Rawlsian "primary good."[287] The arguments have mostly been made in the context of search engines, given their role in distributing information.[288] However, the changing nature, use, and role of social media platforms has made them primary players in the production, dissemination, and consumption of information and in some ways they have replaced traditional public forums and their functions.[289]

[104] There may therefore be room to argue for "neutral" access to information on such platforms, much like fair/neutral access to the Internet in the context of the net neutrality issue.[290] This is so because information

http://www.zephoria.org/thoughts/archives/2010/05/15/facebook-is-a-utility-utilities-get-regulated.html, https://perma.cc/Z9DQ-QSCP, https://perma.cc/SP77-RZ7N.

[285] *See* Lucas D. Introna & Helen Nissenbaum*, Shaping the Web: Why the Politics of Search Engines Matters*, 16 THE INFO. SOC'Y 169, 178-79 (2000). *See also* Jeroen van den Hoven & Emma Rooksby*, Distributive Justice and the Value of Information: A (Broadly) Rawlsian Approach*, INFO. TECH. & MORAL PHIL. 376, 376 (Jeroen van den Hoven & John Weckert eds., 2008).

[286] Introna & Nissenbaum, *supra* note 285, at 178.

[287] *Id.* at 179; *see* van den Hoven & Rooksby, *supra* note 285, at 381.

[288] *See* Introna & Nissenbaum, *supra* note 285, at 179–80.

[289] *See* Napoli & Caplan, *supra* note 265, at 24.

[290] *See id.* at 5–6.

is a *primary* public good, and thus its distribution should not be left up to the market.[291] Just as we do not want broadband providers controlling the Web driven by their economic interests, we do not want information/news being controlled by platforms driven by advertising dollars. The issue becomes more pressing when we bring the idea Zero Rating into the picture.[292] If Facebook becomes the only way to access the Internet (a public good) for free, the arguments for it being a utility become stronger.[293] If it operates like a public utility, there must be some way of ensuring that the public benefit is not pushed in the sidelines.

[105] Net neutrality is not perfectly analogous because it is more focused on barriers to entry for entrepreneurs and small companies, whereas the idea I am suggesting is audience centric (*i.e.* it is the audience that has the right of access to information). What I am suggesting however, is essentially that the FCC has already initiated *neutrality* regulations in the online world, and while that was on the network layer and not the platform layer, the boundaries between layers and their functions are becoming much less clear.[294]

[106] For instance, when Facebook starts being an Internet provider (Zero Rating) and a content provider (Facebook Instant articles), or when cable companies distribute their own content as well, the distinction stops making much sense.

---

[291] *See* Introna & Nissembaum, *supra* note 285, at 176–77.

[292] *See generally* Corynne McSherry et al., *Zero Rating: What It Is and Why You Should Care,* ELECTRONIC FRONTIER FOUNDATION (Feb. 18, 2016), https://www.eff.org/deeplinks/2016/02/zero-rating-what-it-is-why-you-should-care, https://perma.cc/NJ79-P6JF (discussing the potential negative side-effects if a zero rating system were implemented and how public information could be negatively influenced by the zero-rating plan).

[293] *See* Napoli & Caplan, *supra* note 265, at 23–24.

[294] *See id.* at 5–6.

[107] On this note, Frank Pasquale makes a convincing point that clear analogies are falling apart in the new information environment, calling for attention to the true nature and stakes of the dispute.[295] In his words:

> When massive platforms combine the functions of conduits, content providers, and data brokers, analogies from old free expression cases quickly fall apart. Too many discussions of the expressive dimensions of new media are nevertheless moored in murky doctrinal categories, reifications, and inapt historical analogies that do more to obscure than reveal the true stakes of disputes. It is time to think beyond the old categories and to develop a new way of balancing dominant platforms' rights and responsibilities.[296]

[108] In that sense, the analogies are useful to bring out the principles they are based on, but not for categorization purposes. It is on this idea that the analysis is based, because the media access cases discussed above illustrate clearly the First Amendment values the courts were trying to protect.[297] Similarly, the Net Neutrality principle brings out a notion of fair and neutral access to a public good. When Tim Wu first coined the term in 2005, Facebook was in its first years of existence and looked nothing like today's version; there was a much clearer distinction between the platform/application layer and the network layer.[298] Wu's original concept was based on the notion of promoting "fair evolutionary competition" in a privately owned environment, that would promote "a Darwinian competition among every conceivable use of the Internet so

[295] *See* Pasquale, *supra* note 243, at 512.

[296] *Id.*

[297] *See id.*

[298] *See* Tim Wu, *Net Neutrality, Broadband Discrimination*, 2 J. OF TELECOMM.S & HIGH TECH. L. 141, 171 (2003) [hereinafter NET NEUTRALITY].

that the only the best survive."[299] He viewed the role of regulation in such contexts as warranted to ensure "the short-term interests of the owner do not prevent the best products or applications becoming available to end-users."[300]

[109] Viewed this way, the marketplace of ideas clearly qualifies, whether we view the user as a sovereign consumer or a responsible citizen. The government has, in the past, been allowed to intervene in private property when the goal was exposure to free speech and diverse views. In *Pruneyard Shopping Center v. Robins*, the Supreme Court affirmed a California Statute that resulted in shopping mall owners being required to keep their property open for expressive activity.[301] Moreover, the short-term interests of the platform owners are not aligned with this competitive notion, because their profits are higher when users are exposed to content that is familiar and easy for them to consume and go through quickly.[302]

CONCLUSION

[110] What this paper has tried to illustrate is that the current information environment, as produced by social media filtering algorithms that apply to news and social updates indiscriminately, undermines the notion of true autonomous choice in the selection and consumption of content; and threatens the viability of a functioning marketplace of ideas, a prerequisite for citizens in a democratic society to perform their civic duties. The level of behavioral engineering or even "manipulation" that Big Data technologies have enabled, suggests that market powers alone cannot correct the problems, nor can simple disclosure requirements and notices. I

[299] *Id.* at 142.

[300] *Id.*

[301] Pruneyard Shopping Center v. Robins, 447 U.S. 74, 94 (1980).

[302] *See* Wu, *Net Neutrality*, *supra* note 298, at 142–43, 146.

have further suggested that introducing regulation that addresses the nature of the filtering algorithms is desirable and aligned with constitutional values. The precise details of such regulation will have to be a combined effort of the technical and legal community in order for it to be viable. While it would be unrealistic to think that regulatory intervention would perfectly fix the problems, some of which preexisted the technology in question, it is nevertheless a necessary step in the right direction.