Fordham Intellectual Property, Media and Entertainment Law Journal

Volume 27, *Number* 4 *Article* 3

Autonomy Challenges in the Age of Big Data

Sofia Grafanaki*

*Data Elite, sofiagraf@gmail.com

Autonomy Challenges in the Age of Big Data*

Sofia Grafanaki

Abstract

This Article examines how technological advances in the field of "Big Data" challenge meaningful individual autonomy (and by extension democracy), are redefining the process of self-formation and the relationship between self and society, and can cause harm that cannot be addressed under current regulatory frameworks. Adopting a theory of autonomy that includes both the exploration process an individual goes through in order to develop authentic and independent desires that lead to his actions, as well as the independence of the actions and decisions themselves, this Article identifies three distinct categories of autonomy challenges that Big Data technologies present. The first is the increasing rise of lots of "little brothers," putting individuals in a state of constant surveillance, the very knowledge of which undermines individual self-determination. In the governmental context, the idea of always being watched has long been established as a threat to freedom of expression, free speech, "intellectual privacy," and associational freedoms. The discussion does not focus on government surveillance per se, but draws from the same reasoning to illustrate how similar dangers are present even when it is not the government or a single entity behind the surveillance. The second is an algorithmic self-reinforcing loop in every aspect of our lives, as in a world where everything is tracked, the "choices" one is given are based on assumptions about him, and these same "choices" are the ones that determine and become the new assumption, thereby creating a constantly fortified self-fulfilling prophecy. The very structure of the algorithms used is based on statistical models trained to ignore outliers, collect (im)perfect information about the past and use that to recreate the future. This is true both on an individual level and for society more generally. The third is the use of persuasive computing techniques, allowing companies to move beyond simply measuring customer behavior to creating products that are designed with the specific goal of forming new habits. Finally, this Article demonstrates the need for the development of a vocabulary to assess the ethical, political, and sociological values of these algorithms, and for a full set of ethical norms that can lay the foundations of democracy on the web.

KEYWORDS: Autonomy, Big Data

** LL.M, New York University Law School; MBA, Columbia Business School; B.A., Oxford University. The author is Chief Operations Officer of Data Elite, an accelerator and incubator doing seed investments by providing early stage funding and counseling for Big Data start-ups.

Autonomy Challenges in the Age of Big Data

Sofia Grafanaki*

This Article examines how technological advances in the field of "Big Data" challenge meaningful individual autonomy (and by extension democracy), are redefining the process of self-formation and the relationship between self and society, and can cause harm that cannot be addressed under current regulatory frameworks. Adopting a theory of autonomy that includes both the exploration process an individual goes through in order to develop authentic and independent desires that lead to his actions, as well as the independence of the actions and decisions themselves, this Article identifies three distinct categories of autonomy challenges that Big Data technologies present. The first is the increasing rise of lots of "little brothers," putting individuals in a state of constant surveillance, the very knowledge of which undermines individual self-determination. In the governmental context, the idea of always being watched has long been established as a threat to freedom of expression, free speech, "intellectual privacy," and associational freedoms. The discussion does not focus on government surveillance per se, but draws from the same reasoning to illustrate how similar dangers are present even when it is not the government or a single entity behind the surveillance. The second is an algorithmic self-reinforcing loop in every aspect of our lives, as in a world where everything is tracked, the "choices" one is given are based on assumptions about him, and these same "choices" are the ones that determine and become the new assumption, thereby creating a constantly fortified self-fulfilling prophecy. The very structure of the algorithms used is based on statistical models trained to ignore outliers, collect (im)perfect information about the past and use that to recreate the

* LL.M, New York University Law School; MBA, Columbia Business School; B.A., Oxford University. The author is Chief Operations Officer of Data Elite, an accelerator and incubator doing seed investments by providing early stage funding and counseling for Big Data start-ups.

future. This is true both on an individual level and for society more generally. The third is the use of persuasive computing techniques, allowing companies to move beyond simply measuring customer behavior to creating products that are designed with the specific goal of forming new habits. Finally, this Article demonstrates the need for the development of a vocabulary to assess the ethical, political, and sociological values of these algorithms, and for a full set of ethical norms that can lay the foundations of democracy on the web.

INTRODUCTION

In today's Information Age,[1] privacy has taken on a role of utmost significance for freedom and democracy, and one of the main issues of information privacy concerns the power of commercial and governmental entities over individual autonomy and decision-making.[2] Given the vast quantity of personal information that these entities have access to and collect, and in light of technological developments in the field of "Big Data," which allow for the processing of such information in novel ways, there are growing concerns that we are "sleepwalking"[3] into a future of algorithmic regulation, where decisions about individuals and society in general are made by software taking into account thousands of variables not interpretable in human language.[4] Technology writers have gone so far as to talk about "the end of theory," claiming that the scientific method is becoming obsolete.[5] 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."[6] Viewed this way, "[B]ig [D]ata . . . is a way of thinking about knowledge through data and a framework for supporting decision making, rationalizing action, and guiding practice."[7]

---

1 Broadly speaking, the Information Age is a period of history following the Industrial Age where the digital revolution had brought about an economic and social environment based on information and computerization. *See Information Age*, Merriam-Webster, https://www.merriam-webster.com/dictionary/Information%20Age [https://perma.cc/BC36-L4RW] (last visited Apr. 25, 2017).

2 DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 2 (5th ed. 2015).

3 Evgeny Morozov, *The Real Privacy Problem*, MIT TECH. REV. (Oct. 22, 2013), http://www.technologyreview.com/featuredstory/520426/the-real-privacy-problem/ [https://perma.cc/8329-JQ6V].

4 *See* Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1519.

5 *See* Chris Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, WIRED (June 23, 2008, 12:00 PM), http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory [https://perma.cc/87EK-LCPT].

6 *See* 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., 2014).

7 *Id.*

Concerns over automated decision-making are not new to the recent wave of Big Data technologies. Worries that individual activities can be accurately reconstructed through automated processing were expressed two decades ago, noting that personal information was increasingly used to enforce standards of behavior. Information processing was already seen as an essential element to long-term strategies of manipulation "intended to mold and adjust individual conduct," thus making surveillance the order of the day.[8]

New Big Data technologies have accelerated this process exponentially, and when coupled with a changing society that is becoming more exhibitionistic and intrusive, we are faced with eroding privacy expectations.[9] On the one hand, there is unprecedented deliberate "sharing" of personal information, such as in the context of social networks. On the other hand, we are becoming increasingly dependent on the use of applications ("apps") and the Internet of Things,[10] which, in order to be useful, must track, collect, process, and oftentimes disclose intimate details about their users. An individual may give out bits of information in different contexts—each transfer appearing innocuous. However, the danger is created by the aggregation of information, a state of affairs typically created by hundreds of actors over a long period of time.[11] When such data is aggregated, the resulting picture can be very invasive in private life. Its potential uses are vast and unknown, well beyond the scope of marketing and advertising. Such aggregation is already happening, with the whole industry of data brokers focusing on this very practice.

A recent Federal Trade Commission ("FTC") report provided a detailed analysis on the data broker industry. It called for transpa-

---

[8] Spiros Simitis, *Reviewing Privacy in an Information Society*, 135 U. PA. L. REV. 707, 710 (1987).

[9] *See* Anita L. Allen, *Coercing Privacy*, 40 WM. & MARY L. REV. 723, 735 (1999).

[10] The "Internet of Things" refers to the myriad of interconnected devices that create an online infrastructure of information. *See generally* Julie Kantor, *Get Ready: What You Need to Know About the Internet of Things*, HUFFINGTON POST BLOG (Oct. 7, 2016, 4:16 PM), http://www.huffingtonpost.com/julie-kantor/get-ready-what-you-need-t_b_12387194.html [https://perma.cc/L8D8-SHWR].

[11] Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1432 (2001).

rency and accountability, highlighting that many data broker practices fall outside of any specific laws that require the industry to be transparent, to provide consumers with access to data, or take steps to ensure that the data they maintain is accurate.[12] In summary, data brokers collect information about individuals across many sources, aggregate and analyze it, and subsequently share or sell that information, or information derived from it, to companies or government agencies that use it for purposes including targeted advertising and marketing, verifying an individual's identity, providing "people search" services, and detecting fraud.[13] Some data brokers also have a specific line of business as consumer reporting agencies, which provide reports for purposes of credit applications, insurance, employment, or health care.[14] The sources of their information include (a) (federal, state, and local) government sources, such as census responses, voter registration information, motor vehicle and driving records, and court records; (b) publicly available sources, including social media blogs and other information individuals post on the Internet; and (c) commercial data sources, including web browsing histories and transaction-specific data about purchases from retailers and catalog companies or financial services companies.[15] The FTC report further pointed out that in developing their products, data brokers not only use the raw data they obtain from their sources, but also derive additional data, and use the actual and derived data elements to place consumers in categories ("data segments"). This is done by combining data elements to create lists of consumers with similar characteristics and by developing complex models to predict behaviors. Finally, the report highlighted privacy concerns, explaining that because data brokers have no direct relationship with consumers, consumers are often unaware of their existence, let alone the variety of practices in which they engage. Data may change hands many times along the way from source to end product and as a result, even if consumers

---

[12] FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY C-3 (2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf [https://perma.cc/CDE4-ZUSL].

[13] *Id.*

[14] *See id.* at 7–9.

[15] *Id.* at 11.

had full access to their respective profiles, it would be effectively impossible for them to identify the sources of data used and who else has seen their information.[16]

Much of the existing legal framework protecting privacy, often referred to as "privacy self-management," is based on the notions of notice, access, and consent (to the collection, use, and disclosure of personal data) and assumes that such concepts can give individuals control over their personal information.[17] While there is a common notion that users "pay" for online products and services with personal data, there are strong arguments that such an analogy between payment and online data collection is seriously misleading. Unlike functioning markets, in the context of such data collection, individuals do not know the "prices" they are paying for such products.[18] Intellectual property law professor and privacy expert Katherine Strandburg has pointed out that, from the standpoint of each particular information transaction, unlike ordinary sales transactions where individuals can assess the disutility they will incur by turning over a particular amount, in online transactions, individuals will not have enough information to make a reasonable assessment of the "expected disutility" the particular collection will cause them.[19] The information needed relates to *unknown* future uses or misuses of that information by the data recipient or *unknown* others, which may cause *unknown* harms.[20] Further, "payments" are

---

[16] *Id.* at C-3.

[17] *See* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1880 (2013).

[18] Katherine J. Strandburg, *Free Fall: The Online Market's Consumer Preference Disconnect*, 2013 U. CHI. LEGAL F. 95, 96 (2012).

[19] *See id.* at 132.

[20] *See id.* Strandburg further explained:

> First, users lack information about the types of harms that may arise from data collection, the prevalence of those harms, and their costs. Second, users lack detailed and useful information about company practices involving data collection, storage, and use. Third, users lack information about how any given instance of data collection fits into the data about them that is already flowing in the online ecosystem. Without these three types of information, Internet users cannot make meaningful assessments of the marginal expected disutility of any given use of an online product or service. Even if they had the necessary information, bounded capacity for information processing and bounded rationality would interfere with their ability to assess

not obvious to the consumer, as data collection occurs quietly and incrementally and is not apparent unless and until some detectable and traceable potential harm comes to fruition.[21]

Despite growing privacy concerns indicated by several studies, individuals seem willing to give up their privacy in exchange for services without much thought and only seldom adopt privacy protective technologies.[22] These apparent inconsistencies are based on 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.[23] As a result, consent to the collection, use, and disclosure of personal data, even if binding in a legal sense, is often not meaningful, in the sense of providing an individual with real control over their data.[24] 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.[25] In fact, research from the Pew Research Center found that ninety-one percent of American adults "agree" or "strongly agree" that they have lost control over how their information is collected and used by companies.[26]

The result is a heightened threat to individual autonomy because one's capacity and facility for choice requires a degree of freedom from monitoring, scrutiny, interference, and categorization by others. This autonomy that privacy protects has a broad

---

> their expected disutility and compare it to the expected utility of a given online product or service.

Id. at 133.

21 *Id.* at 150.

22 Alessandro Acquisti & Jens Grossklags, *Privacy and Rationality in Individual Decision Making*, 3 IEEE SECURITY & PRIVACY 26, 26 (2005).

23 *See id.* at 27. *See generally* HERBERT A. SIMON, MODELS OF BOUNDED RATIONALITY (MIT Press 1982).

24 *See* Solove, *supra* note 17, at 1880.

25 *See* Strandburg, *supra* note 18, at 150.

26 Mary Madden, *Public Perceptions of Privacy and Security in the Post-Snowden Era*, PEW RES. CTR. (Nov. 12, 2014), http://www.pewinternet.org/2014/11/12/public-privacy-perceptions [https://perma.cc/JW7U-E8VG].

social value as the cornerstone of a democratic society.[27] The argument proceeds as follows: Part I briefly introduces theories of autonomy that are adopted throughout the discussion. Part II addresses the effect Big Data technologies have on individuals from a psychological perspective and how they can change the ways in which individuals learn, act, and express themselves. Part III discusses how the use of algorithms in the Big Data context create self-reinforcing loops that can interfere with self-exploration and limit available resources and choices, both on the individual level and in the social sphere. Part IV describes persuasive computing techniques and the extent to which they can interfere with an individual's decision-making process.

I. INDIVIDUAL AUTONOMY

Scholars have extensively debated the topic of individual autonomy, its conditions, and its value. Thus, the brief discussion that follows cannot be seen as a complete account on the subject. It is simply intended to express the point of view this Article adopts and to frame the context for analyzing the impact Big Data algorithms have on individual autonomy and free choice.

The basic premise adopted is that autonomy concerns not just one's actions, but also the independence and authenticity of the desires (values, emotions, etc.) that move one to act in the first place. This "implies the ability to reflect wholly on oneself, to accept or reject one's values, connections, and self-defining features, and change such elements of one's life at will."[28] In the context of personal autonomy, "there are aspects of himself that the individual does not fully understand but is slowly exploring and shaping as he develops."[29] The autonomy that privacy protects is "vital to the development of individuality and consciousness of individual choice in life . . . . [T]his development of individuality is particular-

---

[27] *See* Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373, 1426 (2000).

[28] John Christman, *Autonomy in Moral and Political Philosophy*, STAN. ENCYCLOPEDIA PHIL. ARCHIVE (Jan. 9, 2015), http://plato.stanford.edu/archives/spr2015/entries/autonomy-moral/ [https://perma.cc/4A2G-J2LG].

[29] DANIEL J. SOLOVE, MARC ROTENBERG & PAUL M. SCHWARTZ, PRIVACY, INFORMATION, AND TECHNOLOGY 37 (2006).

ly important in democratic societies, since qualities of independent thought, diversity of views, and non-conformity are considered desirable traits for individuals."[30] Such independence requires "time for sheltered experimentation and testing of ideas, for preparation and practice in thought and conduct, without fear of ridicule or penalty, and for the opportunity to alter opinions before making them public."[31]

Georgetown Law professor Julie Cohen takes a similar approach, arguing that what we are looking for is *meaningful* autonomy and we must focus on *how* one develops the capacity and ability for choice.[32] To exist, in fact as well as in theory, autonomy must be nurtured.[33] For Cohen, "autonomous individuals do not spring full-blown from the womb." She wrote: "We must learn to process information and to draw our own conclusion about the world around us. We must learn to choose and we must learn something before we can choose anything."[34] Cohen explained how, in a contingent world (referring to society, environment, and circumstance), autonomy requires a zone of relative insulation from outside scrutiny and interference—"a field of operation within which to engage in the conscious construction of self . . . where one can experiment not just with beliefs and associations, but also with every other conceivable type of taste and behavior that expresses and defines self."[35]

This is important not only in the context of individual freedoms, but also as a prerequisite for a democratic society. For Berkeley Law professor and information privacy expert Paul Schwartz, self-determination is a capacity that is embodied and developed through social forms and practices: "The threat to this quality arises when private or government action interferes with a person's control of her reasoning process."[36] The maintenance of a democratic order requires both deliberative democracy and an in-

---

30 *Id.* at 38.

31 *Id.*

32 Cohen, *supra* note 27, at 1426.

33 *Id.*

34 *Id.* at 1424.

35 *Id.* at 1424–25.

36 Paul M. Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1609, 1655 (1999).

dividual capacity for self-determination, and Schwartz remarked that the emerging pattern of information use in cyberspace poses a risk to these two essential values. Schwartz further argued that "perfected surveillance of naked thought's digital expression short-circuits the individual's own process of decision-making."[37] Without a realm of autonomous, unmonitored choice, vital diversity of speech and behavior, as well as constitutionally protected decisions about political and intellectual association, may be chilled.[38]

Digital technologies have challenged notions of autonomy to such an extent that recent scholarship suggests alternative grounds for valuing privacy. Notably, philosophy researcher Tobias Matzner, whose work explores the intersection of politics and technology, proposed a perspective of privacy that moves away from protecting an existing person against a socio-technical background, but rather assumes that we are all inevitably socio-technically "tainted" and focuses on protecting the distance between different appearances (i.e., ways of being a person), each of those contingent to their respective contexts and audiences and the power relations that form their structures.[39] The discussion in Part III returns to this idea of different appearances and the challenges that Matzner identified with the concept of identity management in the digital context. Matzner concluded that privacy moderates and contributes to the ways in which a person comes to being and protects the "freedom to appear and have one's personality negotiated here and now—rather than being determined by all kinds of data."[40]

The perspective adopted in this Article assumes that the subject of autonomy includes the ways in which we become who we are in different contexts, and therefore categorizes challenges to personality formation as challenges to autonomy. On this basis, autonomy is visualized as having two separate stages for every action (or decision) an individual does or does not take. The first is an exploration phase that the individual goes through, seen as both a general exploration of one's true self (whether that is one "real

---

[37] *Id.* at 1656.

[38] *See* Cohen, *supra* note 27, at 1424.

[39] Tobias Matzner, *The Subject of Privacy* 19 (New Sch., Working Paper, 2016) (on file with the Fordham Intellectual Property, Media & Entertainment Law Journal).

[40] *Id.* at 22.

self" or one of the many appearances of the self) and as the exploration that relates to a specific act or decision. The second stage is the act or decision itself. Independence and authenticity are required at both stages for autonomy to be meaningful. The central aim of this Article is to illustrate that when Big-Data technologies are involved, they can interfere at several points and in multiple ways in the course of this two-stage process, with the individuals involved being unconscious of their interference most of the time.

Today, not all our formative and decision-making experiences are occurring online or when we are "connected." One could argue that our online actions/interactions are just one of the many ways in which we learn and decide, and thus are not a substantial threat to our autonomy. But the instances and the complexity in 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. We are at the early stages of a new technological revolution that is evolving at an exponential rather than a linear pace and "will fundamentally alter the way we live, work, and relate to one another."[41] This "Fourth Industrial Revolution" is characterized by a convergence of the digital, physical, and biological spheres.[42]

On the one hand we have "billions of people connected by mobile devices, with unprecedented processing power, storage capacity, and access to knowledge . . . [,]" and on the other, we are witnessing breakthroughs in emerging fields such as artificial intelligence, robotics, the Internet of Things, 3-D printing, synthetic biology, nanotechnology, biotechnology, materials science, energy storage, and quantum computing.[43] Put together, these are creating a "symbiosis between microorganisms, our bodies, the products we consume, and even the buildings we inhabit."[44]

---

41 *See* Klaus Schwab, *The Fourth Industrial Revolution: What It Means, How to Respond*, WORLD ECON. F. (Jan. 14, 2016), https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond [https://perma.cc/2NEY-Y2HP].

42 *Id.*

43 *Id.*

44 *Id.*

With this broader context in mind, Parts II, III, and IV of this Article proceed to discuss different categories of technologies and different types of interference with individual autonomy. The knowledge of constant surveillance discussed in Part II has chilling effects both on the exploration an individual goes through in the process of becoming a person and on the actions that such a person may take. The algorithms used to curate and filter content discussed in Part III can interfere and alter the individual's exploration process. The persuasive computing techniques discussed in Part IV affect the point of decision-making or action.

Such challenges to privacy and autonomy will only intensify as these new information technologies are adopted by an increasing number of the population, and their use becomes more seamless. Data science is at the core of this new connectivity, thus tracking and sharing of personal information is essential. While the technologies addressed in this Article are very basic when placed in the broader context of the technological breakthroughs we are experiencing, they represent the building blocks of more complex emerging technologies. They were chosen as illustrative because their relative simplicity makes it conceptually possible to identify the challenges they pose and the values, politics, and ethics they embody. Once we move to more complex technologies such as artificial intelligence, it is ironically the machines that become "autonomous," taking control of their own learning processes and making it hard even for engineers, let alone social scientists, to identify exactly how they (will) work and assess their impact on our society and our lives.

II. THE "LITTLE BROTHERS"

The risks to individual self-determination in the context of government surveillance have been long-standing topics both in scholarly literature and in law and culture. In current society, not only are these risks similar when the observer is a private entity or a marketer instead of the government, but we are further faced with an additional set of risks that relate to the interpretation of our digital selves. While information gathering by private companies is not thought of as surveillance in the traditional sense, we live in an age

where the two are deeply intertwined.[45] Behavioral marketing is the funding source of the Internet: Well beyond simple cookies, super cookies can track users even in privacy mode;[46] new tracking technologies, such as canvas fingerprinting are impossible to block;[47] data brokers aggregate and analyze information that is collected from and shared with both government and commercial actors;[48] and the Edward Snowden revelations show the extent to which nongovernment information collection not only supports, but can in some instances be the backbone of government surveillance.[49] Apart from the bulk collection of telephone metadata, which some argue still persists but in a different form,[50] the Snowden revelations indicated a much broader practice by the National Security Agency ("NSA"). By piggybacking off commercial tracking technologies designed to serve personalized advertisements ("ads") to consumers, the NSA significantly expanded its surveillance capabilities.[51]

---

[45] Neil M. Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934, 1938 (2013).

[46] Jose Pagliery, *"Super Cookies" Track You, Even in Privacy Mode*, CNNMONEY (Jan. 9, 2015, 10:03 AM), http://money.cnn.com/2015/01/09/technology/security/super-cookies [https://perma.cc/6X8Q-US6K].

[47] Julia Angwin, *Meet the Online Tracking Device that Is Virtually Impossible to Block*, PROPUBLICA (July 21, 2014, 9:00 AM), http://www.propublica.org/article/meet-the-online-tracking-device-that-is-virtually-impossible-to-block [https://perma.cc/4JGH-C7QQ].

[48] *See* FED. TRADE COMM'N, *supra* note 12, at 11–18; Chris Jay Hoofnagle, *Big Brother's Little Helpers: How Choicepoint and Other Commercial Data Brokers Collect, Process, and Package Your Data for Law Enforcement*, 29 N.C. J. INT'L L. & COM. REG. 595, 636–37 (2004).

[49] *See generally* GLENN GREENWALD, NO PLACE TO HIDE: EDWARD SNOWDEN, THE NSA, AND THE U.S. SURVEILLANCE STATE (Metro. Books 2014).

[50] *See* Robert Hackett, *No NSA Phone Spying Has Not Ended*, FORTUNE (Nov. 30, 2015), http://fortune.com/2015/12/01/nsa-phone-bulk-collection-end [https://perma.cc/AJ53-V5ZD]; *NSA Begins New Phone Surveillance Program as Bulk Metadata Collection Ends*, RT (Nov. 28, 2015, 3:19 AM), https://www.rt.com/usa/323806-nsa-new-phone-surveillance [https://perma.cc/2CBF-ANKW].

[51] *See* Ashkan Soltani, Andrea Peterson & Barton Gellman, *NSA Uses Google Cookies to Pinpoint Targets for Hacking*, WASH. POST: SWITCH (Dec. 10, 2013), http://www.washingtonpost.com/blogs/the-switch/wp/2013/12/10/nsa-uses-google-cookies-to-pinpoint-targets-for-hacking/ [https://perma.cc/8GML-Z45R]. The Snowden documents published specifically mentioned the NSA making use of a Google tracking mechanism, the "PREF" cookie, which is assigned to a user's browser when it connects to any Google services or properties, whether direct (such as Search or Maps) or indirect (such as any website with a Google Plus widget). *Id.* The cookie is of course designed to

Further, the very structure of consumer-facing technologies and systems encourages deliberate information sharing on the part of individuals, both because of seductive potential benefits (ranging from price discounts and fast airport security screenings to social status) and a new wave of self-tracking technologies, promising to improve every aspect of our lives. The result is a series of what some scholars have called "surveillant assemblages" that work together as a functional entity by breaking down information about us into discrete data flows, and then reconstructing our "data double," a digital version of us that is not only the subject of marketing practices, but also the determinant of access to resources, services, and power.[52]

Consequently, we can make the following preliminary observations: First, individuals are becoming more aware of the ways that organizations are tracking them and the ways that one can be monitored are becoming more and more intrusive as technology advances. Second, irrespective of who the observer is, information flows between private and public actors weaken and blur the distinction between government surveillance and commercial information collection. Third, individuals deliberately share more and more information about themselves. And fourth, all information combined can be used to make decisions about the individual, ranging from minor marketing decisions to determining access to resources such as welfare, insurance, and credit. This Part focuses on the psychological effects these factors have on individuals before any decision about them is made. Parts III and IV address the consequences of the decisions.

Invoking metaphors often used by privacy scholars, the consequences can be broadly thought of as belonging to two categories,

---

track users in order to serve them with personalized ads, but when used by the NSA, it can even allow for remote exploitation of a users' computers. *Id.* Given the ubiquitous presence of Google services, most web users are likely to have a PREF cookie on their browser, whether they use Google's services directly or not. *Id.* This is not the only commercial tracking mechanism the NSA piggybacks on, the same is done through apps that track their users' locations, for example. *Id.*

[52] Kevin D. Haggerty & Richard V. Ericson, *The Surveillant Assemblage*, 51 BRIT. J. SOC. 605, 605 (2000).

the first being *Orwellian* or *Panoptic* and the second *Kafkaesque*,[53] with both categories undermining individual self-determination and meaningful autonomy. The former category includes chilling effects that pose a threat to civil liberties, such as freedom of expression, free speech, *intellectual privacy*,[54] and associational freedoms, thus constituting a threat to the very concept of democracy. The latter category refers to the state of helplessness, powerlessness, and vulnerability individuals feel when they do not know what information and personal data institutions may have about them and how this information and data may be used.

A. Orwellian Concerns

As people learn that their every keystroke and mouse click is monitored when they are online, George Orwell's "Thought Police" in *Nineteen Eighty Four* is extended to a concept of the "Cyber Thought Police," able to conduct "perfected surveillance of naked thought's digital expression."[55] The Panopticon,[56] the work of English philosopher and social theorist Jeremy Bentham, is an illustration of how surveillance changes the entire landscape in which people act by transforming one's relation to himself and leading to an internalization of social norms that soon is not even perceived as repressive.[57] Without diving into the full spectrum of the dangers of surveillance, suffice to say that chilling effects and

---

[53] *See* Daniel J. Solove, *"I've Got Nothing to Hide" and Other Misunderstandings of Privacy*, 44 SAN DIEGO L. REV. 745, 766 (2007).

[54] *See generally* Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 387 (2008).

[55] Schwartz, *supra* note 36, at 1656.

[56] *See* Solove, *supra* note 11, at 1414–15. *See generally* JANET SEMPLE, BENTHAM'S PRISON: A STUDY OF THE PANOPTICON PENITENTIARY (1993).

[57] *See* MICHEL FOUCAULT, DISCIPLINE AND PUNISH: THE BIRTH OF PRISON 200 (Alan Sheridan trans., Pantheon Books 1st Am. ed. 1977). The Panopticon was a prison designed around a central surveillance tower from which a warden could see into all of the cells, but prisoners had no idea when they were being watched. *Id.* As a result, prisoners had to assume they were always watched and conformed their activities to those desired by the prison staff, even though it was physically impossible for all cells to be watched at all times. *Id.* at 201. In Bentham's words, "[t]o be incessantly under the eyes of an Inspector is to lose in fact the power of doing ill, and almost the very wish[,]" the prison being "a new mode of obtaining power of mind over mind." JEREMY BENTHAM, THE PANOPTICON WRITINGS 31 (Miran Božovič ed., 1995) (1787); Jeremy Bentham, *Panopticon*, *in* 3 OPINIONS OF DIFFERENT AUTHORS UPON THE PUNISHMENT OF DEATH 321, 328 (Basil Montagu ed., 1816).

threats to individual autonomy and self-determination are present without the need for a "Big Brother" or a totalitarian agenda.[58] The threat is simply a product of aggregation of modern practices across different industries amounting to a vast sum of "little brothers" (or "connected apps") driven by Big Data technologies that are increasingly becoming indispensable to organizations. Put differently, the danger is in the cumulative effect that non-trivial instances have over time and in combination, and not necessarily in a specific extreme act or violation.[59] Privacy expert and law professor Daniel Solove, has pointed out that such privacy problems resemble environmental harms, which occur over time through a series of small acts by different actors, and oftentimes gradual pollution from many different sources can be worse than a major spill.[60] Julie Cohen framed the issue as a "modulated society" of "surveillant assemblages," where surveillance is ordinary and signals a seductive appeal—"its ordinariness lending it extraordinary power."[61]

No matter the description used for the current state of "dataveillance,"[62] it results in a slight adjustment to our behavior, both as we explore and develop, and as we act. In terms of our exploration stage, where an individual is learning and developing, privacy scholar and law professor Neil Richards argued that what is at stake is our *intellectual privacy*, referring to protection from surveillance or interference when we are engaged in the processes of generating ideas—meaning when we are thinking, reading, and speaking with confidantes before our ideas are ready for public consumption.[63] Without such protections, fear or embarrassment, judgment, disapproval, or fear of being revealed and exposed interferes with the ways we explore our ideas, what we read or watch, and how we figure out our personal values, politics, and even sexuality.[64]

---

58 *See generally* GEORGE ORWELL, NINETEEN EIGHTY FOUR (Penguin 2013) (1949).

59 Solove, *supra* note 53, at 769.

60 *Id.*

61 Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904, 1916 (2013).

62 *See generally* Roger Clarke, *Information Technology and Dataveillance*, 31 COMM. ASS'N FOR COMPUTER MACHINERY 498 (1988).

63 *See* Richards, *supra* note 45, at 1934; *see also* Richards, *supra* note 54, at 387.

64 *See* Danielle Citron, *Neil Richards on Why Video Privacy Matters*, CONCURRING OPINIONS (Jan. 4, 2012), http://concurringopinions.com/archives/2012/01/neil-richards-on-why-video-privacy-matters.html [https://perma.cc/V5BC-ZHPM].

Such safeguards have a broader social value based on the idea that free minds are the foundation of a free society.[65] Librarians have long recognized confidentiality of library records as a core value, as lack of privacy and confidentiality chills users' choices and therefore suppresses access to ideas.[66] Now that many library functions (such as reading, research, and education) are performed online, a similar set of values is required in order to protect our civil liberties.[67] Citing cognitive psychology research, Julie Cohen made the points that "it is not that people will not learn under conditions of no privacy, but that they will learn differently" and "experience of being watched will constrain the acceptable spectrum of belief and behavior . . . [and] will at the margin incline choices toward the bland and mainstream."[68]

While such concerns and references to Big Brother or the Panopticon may sound abstract and theoretical, the reality of the chilling effects is demonstrated both by a recent paper demonstrating changes in search keywords,[69] and by a recent report by the Pew Research Center on the effects of the Snowden revelations on Americans' behaviors.[70] Among other statements, the report quoted respondents who said they had modified their behavior as explaining: "I [do not] search some things that I might have before . . . it may appear suspicious, even if my reason is pure curiosity," and pointed out that some people have not adopted tools that could make their activities more private because they believe taking such measures could make them appear suspicious or could trigger

---

65 *See* Richards, *supra* note 54, at 404.

66 *See Policy on Confidentiality of Library Records*, AM. LIBR. ASS'N (July 2, 1986), http://www.ala.org/advocacy/intfreedom/statementspols/otherpolicies/policyconfidentiality [https://perma.cc/JZT4-RLNB].

67 *See* Daniel C. Howe & Helen Nissenbaum, *TrackMeNot: Resisting Surveillance in Web Search*, *in* LESSONS FROM THE IDENTITY TRAIL, ANONYMITY PRIVACY AND IDENTITY IN A NETWORKED SOC'Y 417, 417 (Ian Kerr et al. eds., 2009).

68 Cohen, *supra* note 27, at 1425–26.

69 *See* Alex Matthews & Catherine Tucker, Government Surveillance and Internet Search Behavior (Apr. 29, 2015) (unpublished manuscript), http://ssrn.com/abstract=2412564 [https://perma.cc/JC9H-Y7YF].

70 *See* Lee Rainie & Mary Madden, *Americans' Privacy Strategies Post-Snowden*, PEW RES. CTR. (Mar. 16, 2015), http://www.pewinternet.org/2015/03/16/Americans-Privacy-Strategies-Post-Snowden [https://perma.cc/BNQ9-NYQ8].

additional monitoring.[71] The same ideas apply not just to the exploration stage of our autonomous selves that intellectual privacy protects, but can equally affect actions. Recent reports show that writers are overwhelmingly worried about mass surveillance and are engaging in self-censorship, not just in the United States but globally as well.[72] Beyond intellectual privacy, the Pew Research Center report found that a number of people choose not to use privacy enhancing tools, for example, out of fear of raising suspicions.[73] Similarly, one might not seek help for depression or alcoholism out of fear that a potential employer could find out and might not participate in online forums when the topic is sensitive.[74] Whether the data collected is personally identifiable or not seems irrelevant, because even anonymized data can be reidentified when combined with other available data sets.[75] But even without reidentification or a human in the process, individuals are still "reachable."[76] A simple example is that people can still be served with targeted ads that can reveal a lot about them to other users of the same computer.[77]

B. Kafkaesque Concerns

In a law review article, Daniel Solove adopted the metaphor of Kafka's *The Trial*, referring to the state of helplessness, powerlessness, and vulnerability individuals feel when they do not know what information and personal data institutions may have about them and/or how they may be used to make important decisions about them.[78] In such cases, to adjust one's behavior to conform to the

---

[71] *Id.* The report indicated, among other findings, that thirty-four percent of those who are aware of the surveillance programs have taken at least one step to hide or shield their information from the government. *Id.*

[72] *See* PEN AMERICA, GLOBAL CHILLING: THE IMPACT OF MASS SURVEILLANCE ON INTERNATIONAL WRITERS (2015), http://www.pen.org/sites/default/files/globalchilling_2015.pdf [https://perma.cc/4FWL-C97V]; PEN AMERICA, CHILLING EFFECTS: NSA SURVEILLANCE DRIVES U.S. WRITERS TO SELF-CENSOR (2013), http://www.pen.org/sites/default/files/2014-08-01_Full%20Report_Chilling%20Effects%20w%20Color%20cover-UPDATED.pdf [https://perma.cc/7AGQ-C3FP].

[73] Rainie & Madden, *supra* note 70.

[74] *See id.*

[75] *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1743 (2010).

[76] Barocas & Nissenbaum, *supra* note 6, at 45.

[77] *Id.*

[78] Solove, *supra* note 53, at 756.

mainstream will not necessarily make that person feel safe because the nature of data mining in question is predictive and the process of prediction indecipherable.[79] Even if that person conforms to acceptable standards in the present, thus having "nothing to hide," he is still vulnerable to unknown predictions of his future actions that he will not be in a position to disprove, as they have not yet happened.[80] For Solove, the harms consist of those created by bureaucracies, such as "indifference, errors, abuses, frustration, and lack of transparency and accountability," with these problems also affecting social structure by altering the kind of relationships people have with the institutions that make important decisions about their lives.[81]

While talking of Kafka may again seem abstract or sound like an intellectual exercise, the state of anxiety individuals experience in *The Trial* has started to feel a little too familiar. Kate Crawford, a principal researcher at Microsoft Research, in her article *The Anxieties of Big Data*, drew parallels between the lived reality of Big Data and a *surveillant anxiety*: "[T]he fear that all the data we are shedding every day is too revealing of our intimate selves but may also misrepresent us."[82] Drawing on politically active organizations, such as British group Plan C, as well as consumer trends forecasters, such as New York-based K-Hole, Crawford visualized current cultural anxiety as the point of intersection between mass surveillance and mass consumerism.[83] It seems impossible to escape, as every attempt to do so ends up reinforcing it, due to the respective anxiety of the watchers, who live in the fear of not having enough data to be able to derive something meaningful. Put simply, as people seek ways to avoid data collection about them, more intrusive data collection techniques are developed, both by marketers and the government. Illustrative examples of such tech-

---

79 *Id.*

80 *Id.*

81 *Id.* at 766.

82 Kate Crawford, *The Anxieties of Big Data*, NEW INQUIRY (May 30, 2014), http://thenewinquiry.com/essays/the-anxieties-of-big-data/ [https://perma.cc/GPB5-4PKV].

83 *Id.*

niques are facial recognition technologies,[84] used in the context of law enforcement, security, and marketing, as well as new retail experiences, such as Amazon Go, a checkout-free store that uses the same types of technologies as self-driving cars—namely computer vision, sensor fusion, and deep learning—to track customers as they navigate physical space, just as they are tracked in online space.[85] Amazon is not alone in such efforts; it seems that the offline tracking of shoppers is becoming a new retail trend.[86]

For Plan C, anxiety is today's public secret and the dominant effect of the current state of capitalism, which has spread to the whole social field due to "the multi-faceted omnipresent web of *surveillance*."[87] Political ideology aside, the group's thesis is notable for its description of the relationship between this anxiety and surveillance.[88] It spoke of a bureaucratized public space, from which an individual is excluded if he/she does not participate in "deliberate and ostensibly voluntary self-exposure," and where "a widening range of human activity is criminalized on the grounds of risk, security, nuisance, quality of life, or anti-social behavior."[89] In this space, Plan C argued, we are commanded to communicate, yet we are all "co-actors in an infinitely watched perpetual performance, [and] our success in this performance in turn affects every-

---

[84] *See* Natasha Singer, *Never Forgetting a Face*, N.Y. TIMES (May 17, 2014), https://www.nytimes.com/2014/05/18/technology/never-forgetting-a-face.html?src=xps&_r=0 [https://perma.cc/4WBJ-PCSA].

[85] These efforts include observing the path taken to get to a particular item (similar to tracking which links drive traffic) and tracking items that are picked up but returned to the shelves (similar to browsing and clicking on specific products but not buying them). *See* Carly Page, *Amazon Go Lets You Trade Your Privacy for a Cashier-Less Lunch Buying Experience*, INQUIRER (Dec. 5, 2016), https://www.theinquirer.net/inquirer/news/2479078/amazon-go-lets-you-trade-your-privacy-for-a-cashier-less-lunch-buying-experience [https://perma.cc/V4UV-KPEL]; *see also* Carly Page, *Privacy Groups: Amazon Go Takes Invasive Technologies to a 'Whole New Level,'* INQUIRER (Dec. 5, 2016), http://www.theinquirer.net/inquirer/news/2479169/privacy-groups-amazon-go-takes-invasive-technologies-to-a-whole-new-level [https://perma.cc/7VUM-AQPS].

[86] *A New Industry Has Sprung up Selling "Indoor-Location" Services to Retailers*, ECONOMIST (Dec. 24, 2016), http://www.economist.com/news/business/21712163-there-money-be-made-tracking-shoppers-paths-inside-stores-new-industry-has-sprung-up [https://perma.cc/5JE3-JEXN].

[87] *We Are All Very Anxious*, PLAN C (Apr. 4, 2014, 7:52 PM), http://www.weareplanc.org/blog/we-are-all-very-anxious/ [https://perma.cc/SMN3-77BY].

[88] *Id.*

[89] *Id.*

thing from our ability to access human warmth to our ability to access means of subsistence, not just in the form of the wage but also in the form of credit."[90]

In a 2000 article, Kevin Haggerty and Richard Ericson noted a social transformation they called the "disappearance of disappearance."[91] The freedom for self-creation that once came with moving from small communities to big cities where one could be anonymous, was not only hard to find, as it was increasingly difficult to escape social monitoring,[92] but was also showing its dark side. Modern individuals were experiencing an obligation to be free and to find identities with "no stable ground on which to lodge an anchor."[93] Interestingly, in 2013, consumer trends forecaster group K-Hole came up with the term "normcore" to describe the person who seeks to just "blend-in" as a way out of the exhausting effort to be uniquely individual.[94] The group found that "the job of the advanced consumer is managing anxiety, period."[95] This is because, according to K-Hole, "the markers of individuality are so plentiful and regenerate so quickly that [it is] impossible to keep up" with them, and the need to order and narrate our decisions "produces a feeling of trappedness" because, even though the data exists to make us less nervous, "we feel increasingly pressured to do a better job."[96]

What was once seen as the path to personal freedom, they noted, is now making us more isolated as the terms keep getting more specific.[97] In their words: "Today people are born individuals and have to find their communities . . . . Normcore seeks the freedom that comes with non-exclusivity. It finds liberation in being

---

90 *Id.*

91 Haggerty & Ericson, *supra* note 52, at 605.

92 *Id.*

93 *See generally* ZYGMUNT BAUMAN, POSTMODERNITY AND ITS DISCONTENTS (1997).

94 *Youth Mode: A Report on Freedom*, K-HOLE at 36 (Oct. 2013), http://khole.net/issues/youth-mode [https://perma.cc/P2MM-2SLK].

95 *The K-HOLE Brand Anxiety Matrix*, K-HOLE at 7 (Jan. 2013), http://khole.net/issues/03/ [https://perma.cc/Z7QV-YQCB].

96 *Id.*; *Youth Mode: A Report on Freedom*, *supra* note 94, at 20.

97 *Youth Mode: A Report on Freedom*, *supra* note 94, at 36.

nothing special and realizes that adaptability leads to belonging."[98] The new freedom is in the ability to hide in plain sight.

The different approaches described are meant to illustrate how very real these issues are. From authors and philosophers, to legal scholars and social scientists, to political activists, to hip New York consumer trends groups that blend art with advertisement, the theme is the same: The combination of all the little brothers that are tracking individuals—individuals themselves being one of them—is seriously challenging autonomy and self-determination, and is imposing a vision of self that does not "stand out in the data," but "blends in" by being just another data point among millions that has "nothing to hide."[99]

Selfhood and social shaping do not have to be mutually exclusive, but to preserve meaningful autonomy, we need privacy to act as boundary management.[100] Without the necessary insulation for individual self-determination, individuals are left with a life spent almost entirely in the presence of others, thus becoming increasingly shallow and superficial.[101] While the task of policy makers is sometimes perceived as finding the balance between the individual and the needs of society, such formulations are missing the point. Viewed this way, privacy will always lose against national security, efficiency, entrepreneurship, and the progress of knowledge more broadly.[102] But, as several scholars have noted, privacy is also a public value in the sense that individual autonomy is a prerequisite for a democratic society and innovation, and sacrificing it would be harming that society too.[103] In the latter approach, the balancing task involves the needs of society on both sides of the scale.[104] Ul-

---

98 *Id.* at 27, 36.

99 Crawford, *supra* note 82.

100 *Id.*

101 *See* HANNAH ARENDT, THE HUMAN CONDITION 71 (2d ed. 1998) ("A life spent entirely in public, in the presence of others, becomes, as we would say, shallow. While it retains visibility, it loses the quality of rising into sight from some darker ground which must remain hidden if [it is] not to lose its depth in a very real, non-subjective sense . . . .").

102 *Id.*; *see also* Solove, *supra* note 53, at 753.

103 *See* Cohen, *supra* note 27, at 1427; Cohen, *supra* note 61, at 1912; Simitis, *supra* note 8, at 734; Solove, *supra* note 53, at 763. *See generally* PRISCILLA M. REGAN, LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY (2d ed. 2009).

104 *See* Solove, *supra* note 53, at 763.

timately, the same dangers to democracy arise from both the chilling effects of the Orwellian metaphor and from the imprecise state of anxiety experienced in a Kafkaesque society that is combined with consumerism: A society with a reduced range of viewpoints, implying both reduced freedom for democratic participation and responsible citizenship, and reduced stimuli to awaken the human innovative drive.

III. THE SELF-REINFORCING ALGORITHMIC LOOP

This Part sets out to examine the effect Big Data technologies and algorithms have on the individual's decision-making process (the exploration stage of his autonomy)—in other words, *how* he/she got to a decision, belief, or action. As outlined in Part I, *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.[105] In a contingent world, this process inevitably involves a social context, so the developing self has to continuously engage in boundary management between "autonomous selfhood" and the "reality of social shaping."[106] This Part examines these two forces and their interaction when algorithms are involved by discussing the effects of personalization on the individual sphere and meme culture in the public sphere.

At a very basic level, any individual's development (exploration) depends on the kind of opportunities one is given or denied. In the context of Big Data, several academics have expressed concerns over algorithmic regulation and how it can narrow people's life opportunities in discriminatory ways, especially in the context of scoring.[107] Predictive algorithms used in data mining "learn" from the past by analyzing it and taking into account what the algorithms deem as statistically significant.[108] In the case of individuals, this past is formed from assumptions that include as much as the

---

105 *See* Christman, *supra* note 28.

106 *See* Cohen, *supra* note 61, at 1910.

107 *See generally* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014).

108 *See generally* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

respective algorithm can know about someone, meaning one's characteristics, preferences, habits, personality traits, and anything else that is being tracked and can be inferred.[109] Based on this past, the algorithm conducts predictive analyses and determines what choices that individual will be given.[110] In turn, these same "choices" that one was only given based on assumptions about him are the ones that determine and become the new assumption.[111] This is how the self-reinforcing loop is created. In the abstract, it may be difficult to conceptualize and assess the harm, but when put in context, the consequences become clearer.

On the trivial side of the spectrum is the marketing context where getting only ads for certain types of products based on a profile can hardly qualify as harm. That said, when marketing moves from a simple retail context to other spheres of our lives, such as housing, insurance, credit decisions, and career opportunities, consistently being left out of or denied offers because of a profile can start feeling less trivial. Clearer cases of harm involve discrimination that disadvantages legally protected classes, where algorithms can essentially learn and recreate existing biases from the training data they are exposed to, sometimes even without the intention or knowledge of the humans that programmed them.[112] A simple example is an employer devising an automated system to screen applicants based on previous hiring decisions. If the hiring manager in charge of this process previously had systematically disfavored racial minorities or women, the algorithm will "learn" this bias and recreate it by coding it into the system.[113] But even without classic forms of discrimination present, algorithms can create self-fulfilling prophecies that are hard to escape, an example of which can be found in the credit sector where the indication of financial distress will cause borrowers to be profiled as a higher risk, leading to higher rates and more onerous financial obligations, which in turn rein-

---

[109] *See generally* Kosinski et al., *Private Traits and Attributes Are Predictable from Digital Records of Human Behavior*, 110 PROC. NAT'L ACAD. SCI. 5802 (2013).

[110] *See* Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 253 (2013).

[111] *Id.*

[112] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674 (2016).

[113] *Id*. at 682.

force their financial distress and create a higher risk of default.[114] Such instances create "negative spirals" where the algorithm itself is generating the outcome it is supposed to be predicting,[115] creating a negative self-reinforcing loop.

Such instances that interfere with opportunities one is given can have an impact on the way an individual develops and shapes every aspect of his life as well as his view of the world in general. As such, the autonomy violation is easy to conceptualize and the challenge becomes finding the legal route to liability and remedy of the harm. Great examples of such challenges can be found in discrimination laws, as illustrated by Solon Barocas' and Andrew Selbst's work, which pointed out that, in most instances, existing laws fall short of providing a route to address algorithmic bias and discrimination.[116] However, these algorithmic self-reinforcing loops are now present across many spheres of our daily life (e.g., retail contexts, career contexts, credit decisions, insurance, Google search results, news feeds), and in the absence of algorithmic bias or a missed opportunity that can be pointed to, the challenges become are even more fundamental, relating to articulating the actual harm in the first place. The remainder of the Part examines these types of instances, both in the private and public spheres.

A. Personalization: The Individual Self-Reinforcing Loop

At a very abstract level, to become autonomous as individuals, we must learn to choose, and we must first learn something before we can choose anything.[117] We are not born autonomous; our autonomy must be nurtured within a zone of relative insulation from scrutiny.[118] In the Information Age, a big part of this "learning" takes place in a digital-networked context, with online search and news being the most basic learning tool. "Search" is how we obtain the information we use in order to draw conclusions about the world around us,[119] and, in the digital world, these searches are dy-

---

[114] Citron & Pasquale, *supra* note 107, at 18.

[115] Tal Zarsky, *Understanding Discrimination in the Scored Society*, 89 WASH. L. REV. 1375, 1405–06 (2014).

[116] Barocas & Selbst, *supra* note 112, at 675.

[117] Cohen, *supra* note 27, at 1424.

[118] *See id.*

[119] *See id.* at 1374.

namic rather than relying on stable sources.[120] As such, the nature of the Google Search algorithm undeniably influences our perception of the world, and the search results matter to culture, business, and society in general.[121] Several commentators have characterized search engines as the gatekeepers of the web or the new gatekeepers of information.[122] Google's mission from the outset has been to organize the world's information and make it universally accessible and useful.[123] And while it is undoubtedly a for-profit business that, like any publicly traded company, has a responsibility to increase shareholder value, there is also the view that data monopolists, like Google, are in fact threatening the economy, and their ability to block competitors from entering the market is not at all different from that of other monopolies, such as in the oil or railroad industries.[124]

The analysis of search is relevant both in the narrow sense of Google search results and as a more general concept. Other than the specific mechanics of the Google algorithm, most ideas presented apply to any company that is curating and managing any type of content for us. As technology evolves, the very meaning of "search" changes with it. It is no longer limited to opening a browser page and typing keywords. Mobile technologies have already changed the way we interact with information. Commentators describe emerging technologies like augmented reality as the "keyboard and mouse of the future," the "future of interaction," and a technology that will map directly to our intuition by filtering

---

[120] FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 59 (1st ed. 2015).

[121] James Grimmelmann, *The Google Dilemma,* 53 N.Y. L. SCH. L. REV. 939, 950 (2009).

[122] *See* EVGENY MOROZOV, TO SAVE EVERYTHING CLICK HERE: THE FOLLY OF TECHNOLOGICAL SOLUTIONISM 164–65 (PublicAffairs 2013); ELI PARISER, THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU 60–61 (Penguin Books Ltd. 2011). *See generally* Herman Tavani, *Search Engines and Ethics*, STAN. ENCYCLOPEDIA PHIL. ARCHIVE (Aug. 27, 2012), http://plato.stanford.edu/archives/spr2014/entries/ethics-search [https://perma.cc/AX6F-SN8P].

[123] *Our Company*, GOOGLE, http://www.google.com/about/company/ [https://perma.cc/GQT7-353V] (last visited Mar. 3, 2017).

[124] Kira Radinsky, *Data Monopolists Like Google Are Threatening the Economy*, HARV. BUS. REV. (Mar. 2, 2015), https://hbr.org/2015/03/data-monopolists-like-google-are-threatening-the-economy [https://perma.cc/YJ2U-AAX8]; *see also infra* note 195 and accompanying text (discussing Google antitrust action in Europe).

out even more information than the brain already does to engage reality with less disharmony.[125] While we do not live in "the Matrix," the last three decades have taught us that "the future is arriving faster than ever."[126] Regardless of which technology will be the next platform, Google's mission statement of making the world's information universally accessible and *useful* inevitably involves moving from *content management* to *knowledge management.* Turning content into knowledge is key not just for Google, but for any company in the space, and advances in artificial intelligence are critical to this effort.

1. How Google Works

The initial concept behind the PageRank algorithm came from research paper referencing in the scientific community,[127] where it is well known that, when a paper is cited as a reference in other papers, this lends it more credibility; the more citations a paper gets, the more important it is considered.[128] Larry Page applied this concept to web links, creating an algorithm that ranked websites based on how many links pointed to them, as well as the importance of the websites doing the linking by counting links two steps back.[129] This had nothing to do with who the searcher was and what they knew about him; on the contrary, it was an effort to rank search results based on *objective* standards. The individual consumer could therefore trust that when he was conducting a search, the results he was presented with were ranked by some objective standard that he could understand. Interestingly enough, the Google founders were initially opposed to commercializing their search engine, believing

---

[125] Dan Farber, *The Next Big Thing in Tech: Augmented Reality*, CNET (June 7, 2013, 7:09 AM), https://www.cnet.com/news/the-next-big-thing-in-tech-augmented-reality/ [https://perma.cc/X3XG-YNKN].

[126] *See generally* Andy Gstoll, *Tim Cook on the Digital You*, TECHCRUNCH (Oct. 20, 2016), https://techcrunch.com/2016/10/20/what-tim-cook-thinks-the-digital-you-will-do/ [https://perma.cc/9YJT-8G9T].

[127] *See generally* RICHARD L. BRANDT, THE GOOGLE GUYS: INSIDE THE BRILLIANT MINDS OF GOOGLE FOUNDERS LARRY PAGE AND SERGEY BRIN 42 (Penguin 2011).

[128] *Id.* at 42–43.

[129] *Id.* at 43.

that search was too important to be vulnerable to corrupting influences.[130]

Now, the Google algorithm is constantly tweaked and the factors that go into it are far from clear.[131] There are good reasons for this—most importantly, efforts by Google to keep the search results from being manipulated by fake links and by the growing number of search engine optimization ("SEO") businesses.[132] Additionally, the volume of information available is increasing exponentially, and to deliver *correct* results, Google has to find new and *smarter* ways to filter the results.[133] To put this in context, according to 2014 estimates, the data universe was 4.4 zettabytes—i.e., 4.4 trillion gigabytes[134]—and ninety percent of all data in history was created in the last two years.[135] This digital universe is now doubling in size every two years and is estimated to reach 44 zettabytes by 2020.[136]

Without diving deep into the complexity of factors used to filter search results today, suffice to say that two of the most prominent factors are historical search query logs and their corresponding search result clicks.[137] Studies have shown that the historical search information improves search results up to thirty-one percent.[138] According to the monopoly argument, today's search engines can-

---

130 *See id.* The founders specifically argued against advertising-funded search engines in an academic paper. *See* Sergey Brin & Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, 30 COMPUTER NETWORKS & ISDN SYSTEMS 107 (1998).

131 *See infra* notes 147–48 and accompanying text.

132 *See* Brian Dean, *Google's 200 Ranking Factors: The Complete List*, BACKLINKO (Nov. 5, 2016), http://backlinko.com/google-ranking-factors [https://perma.cc/7F76-GJJC].

133 *Id.*

134 STEVE LOHR, DATA-ISM: THE REVOLUTION TRANSFORMING DECISION MAKING, CONSUMER BEHAVIOR AND ALMOST EVERYTHING ELSE 4 (2015) (providing an estimate of the "data universe" from the International Data Corporation).

135 *See id.*

136 "If the digital [u]niverse were represented by the memory in a stack of tablets, . . . by 2020 there would be 6.6 stacks from the Earth to the Moon." EMC DIG. UNIVERSE, THE DIGITAL UNIVERSE OF OPPORTUNITIES: RICH DATA AND THE INCREASING VALUE OF THE INTERNET OF THINGS (2014), https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm [https://perma.cc/5TFN-J42M].

137 *See* Radinsky, *supra* note 124.

138 *See* Eugene Agichtein, Eric Brill & Susan Dumais, *Improving Web Search Ranking by Incorporating User Behavior Information*, PROC. 29TH ANN. INT'L ACM SIGIR CONF. ON RES. & DEV. INFO. RETRIEVAL 19, 19 (2006); Radinsky, *supra* note 124.

not reach high-quality results without this historical user behavior, making it impossible for new players to enter the market even if they have better algorithms because they lack comprehensive records of previous user behavior.[139] As a result, overall industry competitiveness suffers.[140] Additionally, other critics have highlighted that search engines raise not merely technical issues, but also political ones, with embedded values in their design that systematically exclude certain websites and certain types of websites in favor of others, systematically giving prominence to some at the expense of others.[141]

Apart from these general concerns, efforts to improve search through better filtering have resulted in increased *personalization,* which has moved well outside search to other services, as will be discussed below. From the companies' perspectives, the goal of personalization is to better serve the consumer.[142] By definition, the process is a subjective one, with key factors being *context* and *relevance.* Context is the factor that gives rise to clear information privacy concerns, while relevance challenges the core of meaningful autonomy.

If we look at the market forces behind personalization, it becomes obvious that 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 really speaks to his idiosyncratic interests, desires, and needs—thus giving rise to the motto of *relevance* in Silicon Valley.[143] Google wanted to return the perfect answer to its users, but the perfect answer for one person is not perfect for another (for example, a user typing the word panthers could mean the large wild cats or the Carolina team).[144] For Google to understand what a user means when he types a combination of keywords in the search bar (i.e., provide a *relevant* answer), it needs to understand the *context* of the inquiry. In this case, context can

---

139 *See* Radinsky, *supra* note 124.

140 *See id.*

141 Lucas D. Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matter*, 16 INFO. SOC'Y 169, 169 (2000).

142 *See, e.g.*, *Our Products*, GOOGLE, http://www.google.com/about/company/products [https://perma.cc/Q7S9-K76V] (last visited Mar. 3, 2017).

143 PARISER, *supra* note 122, at 21, 24.

144 *Id.* at 33.

only be derived from data, so in order to continue improving its product, Google needs to factor more and more data into the analysis.[145] Given the range of Google's services, an obvious way to improve personalization was to use data about its users collected across all products, which led to a major change in Google's privacy policy in 2012 that allowed it to combine such user data.[146]

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. What exactly is factored in is far from clear, but it goes well beyond obvious factors, like location, to include browsing habits and previous searches, one's social network, what preferences people in their social network have,[147] Facebook likes and shares, authority of Facebook user accounts, Pinterest pins, and Yelp reviews.[148] In short, the more the algorithm knows about a user until the moment of the search (past), the better it can predict what the user wants and will want (in the present and future). As a result, the zone of "insulation from outside scrutiny"[149] that is necessary for autonomy to be nurtured starts to dissipate and the kind of chilling effects described in Part II begin to emerge. Moreover, this *relevance* that Google is trying to provide assumes that users will continue to want the same thing they wanted in the past, and will follow the same behavioral patterns. It also assumes that users will want the same as other people with similar traits. Whether the context of the inquiry is news, politics, or retail, the key to personalization is that the results are *relevant* to the user.

---

[145] MG Siegler, *Marissa Mayer's Next Big Thing: "Contextual Discovery"—Google Results Without Search*, TECHCRUNCH (Dec. 8, 2010), http://techcrunch.com/2010/12/08/googles-next-big-thing/ [https://perma.cc/XTM6-9WNN].

[146] *See Updating Our Privacy Policies*, GOOGLE OFFICIAL BLOG, (Jan. 24, 2012) http://googleblog.blogspot.com/2012/01/updating-our-privacy-policies-and-terms.html [https://perma.cc/BU97-QRD2] ("If [you are] signed in, we may combine information [you have] provided from one service with information from other services. In short, [we will] treat you as a single user across all our products, which will mean a simpler, more intuitive Google experience.").

[147] *See Chapter 7: Personalization & Search Engine Rankings*, SEARCH ENGINE LAND, http://searchengineland.com/guide/seo/personalization-search-engine-rankings [https://perma.cc/982S-3992] (last visited Mar. 3, 2017).

[148] *See* Dean, *supra* note 132.

[149] Cohen, *supra* note 27, at 1424.

Where this can start becoming problematic, is when users do not know what they are looking for. When users do know, personalization can save time and energy by showing them exactly what they wanted with no extra effort. If personalization does not, the only harm is that users end up spending a little more time and effort refining the search or using another source. However, if users do not know what they are looking for, personalization can be the equivalent of someone walking into a library to do research and being given a library catalog where some books are missing or misplaced. The trouble is that users would not even know how this affected their research, if at all. They might have missed out on a crucial book, or on nothing at all.

A good illustration of this is given by Internet activist and author Eli Pariser, who compared two users' search results after typing "Egypt" into the search bar during the 2011 protests.[150] One user's results showed information on the protests, while the other user's results showed nothing of the kind in the first page, but instead showed mostly travel information.[151] Still, one would say that if the latter user was interested in information about the protests, he could simply adjust his search, and even if not, given the importance of the events, he would still see a headline somewhere that would bring the protests to his attention. While this is true, and a personalized Google Search alone may not seem so alarming, we have to look at the broader spectrum to see in how many areas of our lives such personalization is taking place. Taking into account the data market and the practices of data brokers, we see that our online experiences are not a simple series of one-to-one relationships with each service we use, but are increasingly more integrated.[152] Even a quick look at the wide range of Google products[153] can intuitively reveal that they are not a result of a random phenomenon, but part of a well-thought-out strategy. Most importantly, in the information tsunami that we live in, constantly curating the

---

[150] *See* Eli Pariser, *Beware Online "Filter Bubbles,"* TED TALKS (Mar. 2011), http://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles [https://perma.cc/78J9-DY2N].

[151] *Id.*

[152] *See* PARISER, *supra* note 122, at 45.

[153] *See Our Products*, *supra* note 142.

content we consume and adjusting our filters is a lot of work that individuals do not want to or do not have the time to do. Thus, for the most part users passively settle for an "unobjectionable" option given.[154]

2. Beyond Search

As with search, the idea behind personalized news websites—Google News included[155]—is to help individuals find the content that is most relevant to them among the noise.[156] Not surprisingly, one of the factors that determine what one sees, is what the person has clicked on before.[157] Clicking on a specific news item validates that the user is interested in it. Therefore, in the eyes of the algorithm, the user will want to see more of the same content. Other factors may include his location, demographic information, previous searches and browsing habits, social network, and political ideology, if it can be inferred.[158] An individual interested in sports (demonstrated by a history of accessing sports articles) will see more sports related news than someone who has not previously clicked on such news. He then clicks on the sports news he is presented with, the assumption that he wants to see such news is strengthened, and the algorithm continues to populate his feed with sports-related news and information. When we extend a preference like sports to attributes such as political preferences, education level, and income level, we can visualize how seeing only news that match those attributes can begin to narrow not only our perception of the news cycle, but also the range of stimuli available

---

[154] *See* PARISER, *supra* note 122, at 68. Pariser described the theory of least objectionable programming as it originates from researching television viewers' behavior in the 1970s, where it was noticed that, with the increasing number of available channels, people quit channel surfing far more quickly than one might suspect. *Id.* "During most of those thirty-six hours a week (that Americans watch TV), the theory suggests, [we are] not looking for a program in particular . . . . [We are] just looking to be unobjectionably entertained." *Id.*

[155] *See generally About Google News*, GOOGLE NEWS, http://www.google.com/intl/en_us/about_google_news.html [https://perma.cc/6T6P-MEJB] (last visited Mar. 3, 2017).

[156] *See generally* Mike Ananny & Kate Crawford, *Designer or Journalist: Who Shapes the News You Read in Your Favorite Apps?*, NIEMANLAB (Sept. 10, 2014, 10:20 AM), http://www.niemanlab.org/2014/09/designer-or-journalist-who-shapes-the-news-you-read-in-your-favorite-apps/ [https://perma.cc/3WFJ-PHW6].

[157] *See id.*

[158] *See* Kosinski et al., *supra* note 109.

that would lead one to explore and develop interests outside existing ones.

However, we are not there *yet*. Personalization algorithms are not at the point where they work seamlessly in the background, our experiences are not all integrated, and headlines from broadcast media still surround us. What today's technology *is* doing though, is twofold. First, it is making more content available to us. This is because technology both enables more content to be created (by giving rise to phenomena such as citizen journalism,[159] for example) and makes existing content more accessible. Hence, our need for some type of filtering increases. Second, technology is also providing the filtering tools, making it easier for us to stay away from content that is outside our comfort zone—to some degree actively and to some degree passively.[160] The 2016 U.S. presidential election is a testament to that, as the one thing that commentators seem to be in agreement on is that most Americans were not regularly exposed to views outside of their own.[161] The conversation on the election will continue to unfold and it is of course infinitely more complex than filtering algorithms, but the point made is that all these tools that are increasingly integrated in our lives are neither neutral nor inconsequential.

In a different context, when we read books on Amazon's Kindle or use Apple's iBooks app, Amazon and Apple know not only which books we read, but also how fast we read them, which pages take longer to get through, the phrases we highlight, and the words we look up in the dictionary. For Amazon, this information is key to its market dominance, helping it determine what is relevant to us and therefore what other books (or products) to recommend. Given that algorithms are already writing news articles and even

---

[159] *See generally* Kate Bulkley, *The Rise of Citizen Journalism*, GUARDIAN (June 10, 2012, 7:29 PM), https://www.theguardian.com/media/2012/jun/11/rise-of-citizen-journalism [https://perma.cc/7FCW-HTMU].

[160] *See* Danah Boyd, *Why America Is Self-Segregating*, DATA & SOC'Y: POINTS (Jan. 5, 2017), https://points.datasociety.net/why-america-is-self-segregating-d881a39273ab#.6dq0guaos [https://perma.cc/Z3M4-UH55].

[161] *See, e.g.*, Mostafa M. El-Bermawy, *Your Filter Bubble Is Destroying Democracy*, WIRED (Nov. 18, 2016, 5:45 AM), https://www.wired.com/2016/11/filter-bubble-destroying-democracyz [https://perma.cc/4GMY-BV8A].

books,[162] it would not be far-fetched to imagine content being personalized not simply by title, but also by vocabulary to suit our education level, or by language expressions that match our background, leaving little room for expansion beyond where we currently are. While such settings would surely be customizable, the theory of least objectionable programming would again suggest a high likelihood of passive consumption. In other words, Amazon's default setting would become the norm. Like filtering algorithms, default rules are all around us and oftentimes very necessary; we do not want to expend our time and energy for repetitive or mundane choices. But the interests of the "choice architects" who set the defaults are not necessarily aligned with the interests of the readers.[163] In the example, the faster we get through a page, and the more pages we read, translates to more ads, which means more money for the content providers, or more books, ultimately resulting in more profits for Amazon. Some readers may want that, others may not, but defaults will sway their choices nonetheless.

In yet a different context, while Google once claimed that a goal would be to be able to answer questions like "What shall I do tomorrow?," "Which job shall I take?," or "Which college shall I go to?,"[164] LinkedIn has already found a way to answer the last question by launching "Youniversity," a personalized tool that takes into account parameters like one's intended career and job and comes up with suggestions by analyzing the millions of profiles on its database.[165]

---

[162] *See generally* 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/RGP6-JRGE]; Shelley Podolny, Opinion, *If an Algorithm Wrote This, How Would You Even 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 [https://perma.cc/M6RK-L73R]; *infra* Section III.B.2 (discussing the idea of robo-journalism).

[163] For a discussion on "choice architecture," see CASS R. SUNSTEIN, CHOOSING NOT TO CHOOSE (2015).

[164] *See* Caroline Daniel & Maija Palmer, *Google's Goal: To Organise Your Daily Life*, FIN. TIMES (May 22, 2007), http://www.ft.com/intl/cms/s/2/c3e49548-088e-11dc-b11e-000b5df10621.html#axzz3ZGsneXNL [https://perma.cc/5JBV-H7HF]; *Hyper-Personal Search 'Possible,'* BBC NEWS (June 20, 2007), http://news.bbc.co.uk/2/hi/technology/6221256.stm [https://perma.cc/67WZ-7XGF].

[165] *See* Ingrid Lunden, *LinkedIn Flexes Its Search Engine Muscle, Adds College-Finding Tools for Students*, TECHCRUNCH (Oct. 1, 2014), http://techcrunch.com/2014/10/01/

The examples of personalization are endless, but the key takeaway is that "search" is only a building block of technology giants' visions for the future of their companies (and of the world given their dominant position). As in any context, the building block is crucial to the whole endeavor. Google cofounder Sergey Brin once said that "the perfect search engine would be like the mind of God."[166] More recently, and in a less abstract manner, Google CEO Larry Page, tying together projects Google already has under way, articulated a vision that included everything from widespread artificial intelligence to self-driving cars to high-altitude balloons that bring Internet access to the far reaches of the world.[167] He described using computers as still a "clunky" experience, pointing out that computers do not know what we are doing and what we know.[168] To improve search, Google has been moving in this direction with tools like Google Now, a service focused on bringing real-time information to users' mobile devices that tries to anticipate what users need to know before asking and surfaces information to users based on the time of day and the user's current location.[169] Google Glass, the company's wearable display that layered digital information directly over what users see in the real world, provided a tangible preview of what an augmented future could look like (even though it was eventually withdrawn from the market so Google could work on future iterations).[170]

---

linkedin-flexes-its-search-engine-muscle-adds-college-finding-tools-for-students/ [https://perma.cc/NHW7-PFLK]; Dan Tynan, *This Is How LinkedIn Can Help You Pick the Right College for Your Kids*, YAHOO! (Feb. 2, 2015), https://www.yahoo.com/tech/how-linkedin-can-help-you-pick-the-right-college-109825950084.html [https://perma.cc/586L-JYEL].

166 Charles H. Ferguson, *What's Next for Google*, MIT TECH. REV. (Jan. 1, 2005), http://www.technologyreview.com/featuredstory/403532/whats-next-for-google/ [https://perma.cc/63Q5-FMUS].

167 *See* Marcus Wohlsen, *Larry Page Lays Out His Plan for Your Future*, WIRED (Mar. 19, 2014, 7:17 PM), http://www.wired.com/2014/03/larry-page-using-google-build-future-well-living [https://perma.cc/JGL6-9PK5].

168 *Id.*

169 David Pierce, *Google Now Expands Its Reach, Integrates 70 More Services*, WIRED (Apr. 28, 2015, 2:00 PM), http://www.wired.com/2015/04/google-now-app-integrations/ [https://perma.cc/Q54S-SRRX].

170 *See* Nick Bilton, *Why Google Glass Broke*, N.Y. TIMES (Feb. 4, 2015), https://www.nytimes.com/2015/02/05/style/why-google-glass-broke.html [https://perma.cc/WJ3W-3KRD].

But, at the core of improving search, is Google's aggressive pursuit of artificial intelligence ("AI"), illustrated by the hire of famous futurist inventor Ray Kurzweil as Director of Engineering.[171] Kurzweil now heads up a team "developing machine intelligence and natural language understanding."[172] As the futurist author and inventor explained in an interview, Kurzweil and his team work on predicting on a "semantically deep level what you are interested in, not just the topic . . . [but] the specific questions and concerns you have."[173] He envisions that, some years from now, "the majority of search queries will be answered without you actually asking . . . [this artificial mind] will just know this is something that [you are] going to want to see."[174] Put differently, Kurzweil's team aims to reverse the concept of "search" completely. Currently, people use search engines to better understand information; in Kurzweil's future, search engines will use Big Data to better understand people.[175]

Regardless of Kurzweil's futuristic vision, AI is already prevalent, as evident in drones, autonomous cars, virtual assistants, and financial technology software. The recently announced Google Assistant is the next generation of Google Now, and possesses deeper artificial intelligence capabilities.[176] The company's newest hardware, Google Pixel and Google Home are designed to be vessels or "Trojan Horses" for the company to deliver the rich AI experience it has been preparing for, without relying on underlying technology

---

[171] *See generally Ray Kurzweil Biography*, KURZWEIL ACCELERATING INTELLIGENCE, http://www.kurzweilai.net/ray-kurzweil-biography [https://perma.cc/7C3B-86X7] (last visited Mar. 3, 2017).

[172] *Id.*

[173] *See* David J. Hill, *Exclusive Interview with Ray Kurzweil on Future AI Project at Google*, SINGULARITYHUB (Jan. 10, 2013), http://singularityhub.com/2013/01/10/exclusive-interview-with-ray-kurzweil-on-future-ai-project-at-google/ [https://perma.cc/25JR-ETXN].

[174] *Id.*

[175] Ian Kerr & Jessica Earle, *Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy*, 66 STAN. L. REV. ONLINE 65, 65–66 (2013).

[176] Sarah Jacobsson Purewal, *The Difference Between Google Now and Google Assistant,* CNET (Oct. 4, 2016, 12:14 PM), https://www.cnet.com/how-to/the-difference-between-google-now-and-google-assistant/ [https://perma.cc/CF94-6SBP].

built by others.[177] Google Search is already being transformed by the use of a deep learning system called RankBrain in a small percentage of search queries.[178] The same type of deep learning based on neural networks ("networks of hardware and software that mimic the web neurons of the human brain") is also what Google's AlphaGo is based on—the computer system that in March 2016 defeated the Korean grand master of the 2,500-year-old game Go, a game exponentially more complicated than chess.[179] The future of AI and its limitations are beyond the scope of this Article, but suffice to say that at the core of modern AI based on neural nets is software designed to "learn on their own" in ways that engineers can find challenging, if not impossible, to trace.[180]

3. Observations for the Individual

Even if we are in the age of AI, we still consume broadcast media, and personalization algorithms are far from perfect, so they are not yet in a position to take over our *exploration* of the world and of ourselves. But what *is* true today, is that the information we are flooded with and we somehow have to consume keeps increasing, so we inevitably need ways to filter it. We 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. "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.[181] Harvard Law professor and

---

[177] Margaret Rhodes, *All That New Google Hardware? It's a Trojan Horse for AI*, WIRED (Oct. 5, 2016, 3:24 PM), https://www.wired.com/2016/10/new-google-hardware-trojan-horse-ai/ [https://perma.cc/FQE9-2NS7].

[178] Cade Metz, *AI Is Transforming Google Search. The Rest of the Web Is Next*, WIRED (Feb. 4, 2016, 7:00 AM), https://www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/ [https://perma.cc/R9XR-JE8D].

[179] Cade Metz, *Google's AI Wins Fifth and Final Game Against Go Genius Lee Sedol,* WIRED (Mar. 15, 2016, 3:01 AM), https://www.wired.com/2016/03/googles-ai-wins-fifth-final-game-go-genius-lee-sedol/ [https://perma.cc/7NLG-5UAA].

[180] *See id.*; *see also* Cade Metz, *What If Computers Become Smarter than Humans?*, KNOWLEDGE@WHARTON (Nov. 22, 2016), http://knowledge.wharton.upenn.edu/article/will-superhuman-artificial-intelligence-turn-us-paper-clips [https://perma.cc/XJB2-XTGZ].

[181] *See* SUNSTEIN, *supra* note 163, at 5–6.

author Cass Sunstein, in his recent book, provided an in-depth analysis of the value of choice, and whether active choices or defaults are best, including why and when.[182] In Sunstein's view, the rise of personalized default rules is a blessing that can contribute to human freedom, if used in the right way.[183] 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 pointed 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.[184] In those cases, where the concern becomes the "risk of manipulation, compromising human agency and even dignity," he found that it is imperative for the default rule to be made public.[185] But Big Data algorithms make that increasingly hard. Big Data technologies have the power to not just influence, but also *to determine* who we are. 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.

If we can tell anything from the intention of the companies that shape our online experience, it is that they are trying to create a complete personalized experience of the web, and by extension the world, for each of us. Our new connectivity[186] means that being online goes far beyond sitting in front of a browser. Whether Ray Kurzweil's singularity is near or not,[187] some version of a personalized experience of the world is, and having it set up based just on relevance seems unsatisfactory and incomplete. Google is not the only company in the space; all the big players are in the race for the best version of an intelligent assistant. Apple has Siri,[188] Microsoft

---

182 *See generally id.*

183 *Id*. at 16–17.

184 *Id.* at 48.

185 *Id.* at 51.

186 *See* Schwab, *supra* note 41.

187 *See* RAY KURZWEIL, SINGULARITY IS NEAR (Penguin Books 2006) (discussing the law of accelerating returns and predicting an exponential increase in technologies that will lead to a technological singularity in the year 2045, where intelligent machines would be capable of recursive self-improvement and progress will be so rapid that humans will not be able to comprehend it).

188 *See Siri*, APPLE, https://www.apple.com/ios/siri/ [https://perma.cc/C3TH-EC59] (last visited Apr. 25, 2017).

has Cortana,[189] Facebook has M,[190] and Amazon has Alexa.[191] The winner will most likely depend on which company can create the most seamless experience across devices and platforms. In other words, the key is the aggregation of personal information.

First off, in the case of Google, we can immediately observe potential conflicts of interest. Google has begun to embrace "paid inclusion" in some of its products, raising a whole new set of concerns.[192] Paid inclusion aside, the incentives of the company may not always be aligned with those of its users. Because users pay companies like Google with their attention and their data, which the companies then convert to advertising revenue, Google's incentive is to keep users "locked-in" to its services in order to keep collecting information, even if competitors may offer better products.[193] Such efforts are also present in Google's new product development in an attempt to harness the momentum that is moving away from desktop search and direct it to other products that the company can use as platforms for its advertising business.[194] This would seem like a simple rule of business, but for the fact that Google is also the way that users find potentially competing products, raising concerns about some of its practices. Such concerns are illustrated by the recent antitrust actions against Google in Europe.[195] Further, as long as results are *relevant*, users can find some

---

[189] *See Cortana*, MICROSOFT, https://www.microsoft.com/en-us/mobile/experiences/cortana [https://perma.cc/S9QC-UKCV] (last visited Apr. 25, 2017).

[190] *See* Jessi Hempel, *Facebook Launches M, Its Bold Answer to Siri and Cortana*, WIRED (Aug. 26, 2015, 1:00 PM), https://www.wired.com/2015/08/facebook-launches-m-new-kind-virtual-assistant/ [https://perma.cc/GQC3-5G6W].

[191] *See Echo & Alexa Devices*, AMAZON, https://www.amazon.com/echo-superbowl-commercial/b?ie=UTF8&node=9818047011 [https://perma.cc/L3B6-RGST] (last visited Apr. 25, 2017).

[192] *See* Danny Sullivan, *Google's Broken Promises & Who's Running the Search Engine?*, MARKETING LAND (Nov. 17, 2013, 9:00 AM), http://marketingland.com/google-broken-promises-65121 [https://perma.cc/8KGL-LWRK]; Danny Sullivan, *Once Deemed Evil, Google Now Embraces "Paid Inclusion,"* MARKETING LAND (May 30, 2012, 9:15 AM), http://marketingland.com/once-deemed-evil-google-now-embraces-paid-inclusion-13138 [https://perma.cc/6KQY-F383].

[193] *See* PARISER, *supra* note 122, at 40–41.

[194] *See* Wohlsen, *supra* note 167.

[195] *See* James Kanter & Mark Scott, *Europe Challenges Google, Seeing Violations of Its Antitrust Law*, N.Y. TIMES (Apr. 15, 2015), http://www.nytimes.com/2015/04/16/business/international/european-union-google-antitrust-case.html?_r=0 [https://

type of unobjectionable answer and advertisers can find likely customers who will click on the ads and increase revenues for Google or Facebook. In this case, Google's business model based on highly targeted advertising can conflict with its role as gatekeeper of information.

The conflict here is based on an important assumption: the ideology of the Internet as a public good and democratic medium. Lucas Introna and Helen Nissenbaum conducted an elaborate analysis on the reasons for conceiving the Internet as a special kind of public place,[196] and pointed out that the Internet fulfills some of the functions of traditional public spaces (like museums, parks, beaches, and schools) by serving as a medium for artistic expression, space for recreation, a place for exhibiting items of historical and cultural importance, and a place with the ability to educate the public. But most importantly, the Internet functions as a conveyor of information, with access to such being construed as a Rawlsian "primary good."[197]

Using a different approach to illustrate the conflict, law professor Frank Pasquale made an interesting argument that, as algorithmic authorities get to know us better, at some point personalization becomes a relationship mutual enough to trigger the classic duties of professional advisers.[198] The argument suggests that similarly to doctors and lawyers, who are subject to malpractice lawsuits when they fail to meet a certain standard of care, online service providers, acting as information fiduciaries, need to take on some responsibility for ordering Internet choices responsibly, and must consid-

---

perma.cc/9JAZ-8TLA]; Arjun Kharpal, *Google Could Face 3 Separate Antitrust Cases: EU Competition Chief*, CNBC (July 16, 2016, 7:13 AM), http://www.cnbc.com/2016/07/19/google-could-face-3-separate-antitrust-cases-eu-competition-chief.html [https://perma.cc/E24N-VHGX].

196 *See* Introna & Nissenbaum, *supra* note 141, at 178.

197 *Id.*; *see also* Jeroen van den Hoven & Emma Rooksby, *Distributive Justice and the Value of Information: A (Broadly) Rawlsian Approach*, *in* INFORMATION TECHNOLOGY AND MORAL PHILOSOPHY 376, 379–81 (Jeroen van den Hoven & John Weckert eds., 2008).

198 *See* PASQUALE, *supra* note 120, at 168.

er whether their information practices create a conflict of interest and act accordingly.[199]

Second, building on the theory of information as a primary good, the notion of relevance is called into question. As a default rule, relevance is certainly efficient from the companies' perspectives, as it aligns very well with profit-maximizing goals. But using relevance as a driver of personalization also means that we end up in self-reinforcing loops, where the options we see are put through a filter that ignores outliers and assumes that "what has been is what will be."[200] When stimuli are tailored to play to existing inclinations, our choices become narrower without us even being aware of it most of the time. This does not mean we cannot get out of the loop if we *actively* want to. It would, however, require what has been called an "effort tax," and research has consistently shown that the "power of inertia" takes over in most cases.[201] 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. 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 choose them.[202]

But one of the most defining human characteristics is that we are *unpredictably* individual and that is what our freedom is based on.[203] This is ignored when efficiency becomes the highest of val-

---

[199] *Id.*; *see also* Jack M. Balkin, *Information Fiduciaries in the Digital Age*, BALKINIZATION (Mar. 5, 2014), http://balkin.blogspot.com/2014/03/information-fiduciaries-in-digital-age.html [https://perma.cc/QU45-GBPY].

[200] Tene & Polonetsky, *supra* note 110, at 254.

[201] SUNSTEIN, *supra* note 163, at 34–37.

[202] *See* Cohen, *supra* note 61, at 1917.

[203] *See* James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L. J. 1151, 1181 (2004). Whitman discussed the right of inviolate personality in German literature:

> [T]o be free was to exercise free will, and the defining characteristics of creatures with free will was that they were unpredictably individual, creatures whom no science of mechanics or biology could ever capture in their full richness. For Germans who thought of things this way, the purpose of 'freedom' was to allow each individual

ues, leading to an appearance of free choice at 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. While some version of a filter is inevitably necessary, relevance is surely not the only way such a filter can be construed. For Eli Pariser, relevance needs to be balanced with what is important, challenging, and uncomfortable, or a different point of view.[204] Without such balance, the algorithms treat humans as constant and predictable, and view human traits like imperfection, ambiguity, disorder, and the opportunity to err as vices to be eliminated,[205] thus leaving very little room for personal growth. More broadly, democratic participation and responsible citizenship requires a certain amount of discomfort in order to "motivate citizens to pursue improvements in the realization of political and social ideals."[206]

Third, even based on relevance alone, it is not clear which version of ourselves forms this loop. Pariser rightly pointed out that the Google version of us is very different than the Facebook version.[207] The former can be seen as the present, real, raw, or even secret and dark version of ourselves, while the latter can be our aspirational, superficial, or public performance version.[208] There are several public performance versions, as one's Facebook version is also different from the Snapchat, LinkedIn, Instagram, or Match.com versions. Regardless of the version, should the algorithm not also take into account its own influence, both as a filter and as a cause of chilling effects? If our "data double" is constructed by isolated information flows about us that are then put back together,[209] different information flows can create completely different realities for individuals.

> fully to realize his potential as an individual: to give full expression to his peculiar capacities and powers.

Id.

204 *See* Pariser, *supra* note 150.

205 *See* MOROZOV, *supra* note 122, at 158–59.

206 *See* Cohen, *supra* note 61, at 1918.

207 *See* PARISER, *supra* note 122, at 113–15.

208 *See id.*

209 *See* Haggerty & Ericson, *supra* note 52, at 611–14.

The idea of a public performance self and of different versions or appearances of an individual is not a concept new to the digital world. Social theorists like Erving Goffman have talked of *impression management*, meaning that human interactions are like theatrical performances where, depending on the context, an individual creates a different impression of himself.[210] The individual is viewed as an actor wearing different masks in different social contexts.[211] More recently, communications researchers Alice Marwick and Danah Boyd have talked about the "context collapse" in the digital world and the more complicated "networked privacy," based on a notion of *identity management* similar to Goffman's *impression management*.[212] Tobias Matzer discussed these concepts in the context of relative privacy, and illustrated that such notions of privacy presuppose a space where the individual can, without interference, take off the mask and contemplate, decide, and control which mask to put on next, if any.[213] Based on this analysis, identity management requires the zone of insulation from outside scrutiny and interference that Part II of this Article identifies as necessary for meaningful autonomy. Using the same analogy, personalization algorithms are not only eliminating the required safe space, but also taking over the management process without us knowing which mask we are wearing, or if we are wearing a combination of elements from different masks. More fundamentally, it is hard to even define what this safe space looks like. The "back stage" where we take off the mask can no longer be seen as the offline world. A lot of the functions of the back stage are now performed by or through Google, but Google itself is also one of the "front stages."

Finally, regardless of which version of ourselves the algorithm settles on, and whether such version is incomplete, static, or erroneous, when we get to a place where we all experience the world

---

[210] *See generally* ERVING GOFFMAN, THE PRESENTATION OF SELF IN EVERYDAY LIFE (1956).

[211] *Id.* at 150–51.

[212] Alice E. Marwick & Danah Boyd, *Networked Privacy: How Teenagers Negotiate Context in Social Media*, 16 NEW MEDIA & SOC'Y 1051, 1054–56 (2014); *see also* Danah Boyd, *How Context Collapse Was Coined: My Recollection*, ZEPHORIA (Dec. 8, 2013), http://www.zephoria.org/thoughts/archives/2013/12/08/coining-context-collapse.html [https://perma.cc/Z7UZ-9VFG].

[213] *See* Matzner, *supra* note 39, at 5–6.

differently, we start having fewer topics for public discussion. Concerns are already expressed about how personalized campaign messages affect our democracy.[214] Beyond elections, the press has long held a vital place in a functioning democracy, by acting as the fourth estate, i.e., as a watchdog on the government.[215] By informing the public of government actions, the press stimulates dialogue and debate, and the government is held accountable.[216] But this role of the press would be compromised if the political discussion were scattered in silos. Even further, Princeton University sociologist Matthew Salganik pointed out that "if we view the role of cultural products as giving us something to talk about, then the most important thing might be that everyone sees the same thing and not what that thing is."[217] Without some common basis for discussion, our ideas are not challenged and cannot evolve, whatever the context.

B. Meme Culture: The Self-Reinforcing Loop in Society

Turning to the social shaping that partakes in our construction of self, we can observe on a broader scale that the very structure of these algorithms is encouraging a culture that technology commentators have coined as *memefication*, defined as "the tendency to assess everything in terms of how the intended audience is likely to react according to what is known about the audience."[218] This is different from simple popular culture, where trends arise naturally and organically (at least in theory). Products like Google's auto complete,[219] Twitter Trends,[220] Facebook's News Feed,[221] and

---

[214] *See generally* Ira S. Rubinstein, *Voter Privacy in the Age of Big Data*, 2014 WIS. L. REV. 861; Zeynep Tufekci, Opinion, *Beware the Smart Campaign*, N.Y. TIMES (Nov. 16, 2012), http://www.nytimes.com/2012/11/17/opinion/beware-the-big-data-campaign.html [https://perma.cc/6P4D-A5EL].

[215] PIPPA NORRIS, DRIVING DEMOCRACY: DO POWER-SHARING INSTITUTIONS WORK? 189 (2008).

[216] *Id.*

[217] Carl Bialik, *Look at This Article. It's One of Our Most Popular*, WALL ST. J. (May 20, 2009 12:01 AM), http://www.wsj.com/articles/SB124277816017037275 [https://perma.cc/QC84-YNYJ].

[218] MOROZOV, *supra* note 122, at 159.

[219] *Search Using Autocomplete*, GOOGLE, https://support.google.com/websearch/answer/106230?hl=en [https://perma.cc/D5Z4-CNJA] (last visited Apr. 25, 2017).

[220] *FAQs About Trends on Twitter*, TWITTER SUPPORT, https://support.twitter.com/articles/101125 [https://perma.cc/G3K4-T269] (last visited Apr. 25, 2017).

YouTube's Trending videos[222] are, by design, promoting content with the potential of becoming an online hit (i.e., more clicks and more advertising revenue) and, at the same time, playing an increasingly important role in how we navigate information and culture.

1. Social Media

Twitter Trends is a feature that decides which topics are "trending" across the platform.[223] Once a topic receives this status, it automatically attracts even more attention and can flow in national and global conversations.[224] In order to do this, Twitter engineers must make some assumptions about what aspects of the public discussion constitute a trend, decide how these aspects are to be measured, and, after measuring them, feed them back to the public.[225] These assumptions are far from clear, and the obscurity is justified by the effort to keep those who want to game the system in the dark. Communications professor Tarleton Gillespie investigated this in an attempt to understand why certain topics seemed to be censored, and pointed out that assumptions include whether the topic is new to Twitter or has trended before, whether the topic is spiking or has gradual growth, and whether the discussion occurs within the boundaries of "clusters" or spans across such borders (a factor that also makes an assumption as to what constitutes a cluster by taking into account, for example, region, demographics, and whether people follow each other).[226] Gillespie also pointed out that this focus on clusters illustrates that the algorithm considers

---

[221] 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/N62W-GFVQ].

[222] Jamieson Cox, *YouTube Is Making It Easier to Find Viral Videos*, VERGE (Dec. 9, 2015), http://www.theverge.com/2015/12/9/9881106/youtube-trending-tab-viral-videos [https://perma.cc/93D6-UMS6].

[223] *FAQs About Trends on Twitter*, *supra* note 220.

[224] *See id.*

[225] MOROZOV, *supra* note 122, at 158.

[226] *See* Tarleton Gillespie, *Can an Algorithm Be Wrong? Twitter Trends, the Specter of Censorship, and Our Faith in the Algorithms Around Us*, CULTURE DIGITALLY (Oct. 19, 2011), http://culturedigitally.org/2011/10/can-an-algorithm-be-wrong [https://perma.cc/8A27-REA8]. This received a lot of attention when tags like "#occupywallstreet" and "#wikileaks" never made it into the trends, instigating a discussion on whether the factors taken into account have a political nature as well. *Id.*

breadth more important than depth.[227] With the same number of users talking about it, a topic is considered more worthy of the trend status if it is mentioned briefly across different clusters than if it is discussed intensely within the same cluster.[228] While one may agree or disagree with such an assumption about what *should* constitute a trend, what again becomes clear is that the algorithms in play are far from neutral and objective.

Twitter Trends is a good illustrative example because it is a relatively visible tool in which algorithms become curators.[229] Other tools and instances are much less visible. What is notable however is that user behavior on any social media platform is to a very large extent structured by the way the platform is designed; each platform has a certain architecture and certain limitations that can determine how users behave and interact.[230] In the case of Twitter, for example, the character limitation means messages are short and turnover is rapid.[231] The simplicity of the interface makes it suitable for mobile devices and, thus, the platform of choice for users on the move or in high-tension situations (like demonstrations).[232] The lack of a requirement for a mutual relationship of users (follow versus friends) also allows for connected clusters that would otherwise not interact.[233] Facebook has different characteristics, resulting in different content surfacing as more popular, but for both Facebook and Twitter, their respective architectures create inherent limitations in what conclusions can be drawn from the data.[234] For many, social media platforms are in some ways replacing publishers and broadcasters as cultural gatekeepers, and the methods

---

227 *See id.*

228 *See id.*

229 *Id.*

230 *See* Zeynep Tufekci, *Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls*, 2014 PROC. 8TH INT'L AAAI CONF. ON WEBLOGS & SOC. MEDIA 505, 510.

231 *Id.* at 507.

232 *Id.*

233 *Id.*

234 *See id.* In the case of Facebook, for example, how many people "like" something depends largely on how many people saw it. For Twitter, a challenge is that retweets are hard to track and can thus reduce visibility of a topic. *Id.* at 510.

they use have very different dynamics that we are only just beginning to understand and unpack.[235]

The issue of fake news in the 2016 U.S. presidential election brought to the surface a big debate about 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.[236] This discussion is beyond the scope of this Article, but the fact that it is taking place in a broader context is very encouraging.

2. Journalism

In the case of online journalism, the issues are equally complex even without personalization. The obvious part is that, as was the case with personalized news, user clicks play a big role in the way users experience news. In this context, the key factor switches from relevance to *popularity*. When a user clicks on a certain news item, he is taken to a new page with different ads, and more ads equal more revenue for the website. News items or opinions that do not get clicked on as much as others get less visibility, and non-mainstream topics and minority opinions are increasingly pushed into the shadows. The result is an information environment where what an individual sees is not only limited by his own previous choices and existing inclinations as explained above,[237] but the environment in which he made these choices in the first place is already limited by the preferences of the majority.

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. Media organizations are predominantly for-profit com-

---

[235] *See* 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/4LPC-YECT].

[236] *See generally* David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 42 LOY. L.A. L. REV. 373 (2010).

[237] *See* discussion *supra* Part III.

panies with commercial pressures whose business model has always been partly based on advertising. Some see this as a very old problem, based on the simple fact that the entity that helps deliver the news is not the same as the creator of news, and thus their interests and understanding of what they should be doing are not the same.[238]

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 a television station—had some breathing room to use a moral compass in determining what becomes news. There was room for the editor who feels it is his responsibility to instigate public debate and deliberate thinking, for the journalist who wants to educate the public on a certain subject, and for the outlier who thinks differently. Editors had the flexibility to promote content that would sell in order to fund content that their writers wanted to write about. Companies could not know exactly how many eyeballs saw their ads, and broadcasting entities did not know precisely which articles were read, by how many readers, and how quickly. This previous state of the industry is well captured by a quote from John Wanamaker, who famously said: "Half the money I spend on advertising in wasted; the trouble is I [do not] know which half."[239] We are now living in a time in which companies know exactly which half is wasted and will not spend marketing money for ads on content that is not very popular (or clickable). 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 will find it increasingly hard to convince shareholders that minority views, which bring no advertising revenue, are still necessary content.

Social media platforms add an additional layer of complexity because of their increasing role in how news is circulated and consumed. With significant traffic on news websites coming from so-

---

[238] *See* O'Donovan, *supra* note 235.

[239] MARILYN ROSS & SUE COLLIER, THE COMPLETE GUIDE TO SELF-PUBLISHING 344 (5th ed. 2010).

cial media platforms,[240] editorial decisions may now include calculations that address the ways in which specific platforms filter content. Much like the Google search algorithm, however, these algorithms are constantly tweaked[241]—a fact that in some ways offsets the dangers of gaming the system but also implies a complete lack of transparency. According to a 2016 survey by Pew Research Center, sixty-two percent of U.S. adults get news on social media, and eighteen percent do so often.[242] Social media news consumers still get news from a variety of other sources, and to a fairly consistent degree, according to the study, but as compared to a 2013, there is a notable increase in news consumption on Facebook, Instagram, and LinkedIn.[243] Other questions addressed included which social media platforms have the largest portion of users getting news from the platform, how many get news on multiple social media platforms, and to what degree these news consumers are seeking online news out versus happening upon it while doing other things.[244] These types of issues are now putting pressure on reporters to write "clickbait" articles that "pander to users' worst impulses."[245]

Media scholar Chris Anderson pointed out another way in which we are facing a problem: There is a fundamental transformation in journalists' understanding of their audiences. [246] Anderson referred 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

---

[240] *See* Mathew Ingram, *Facebook Has Taken Over from Google as a Traffic Source for News*, FORTUNE (Aug. 18, 2015), http://fortune.com/2015/08/18/facebook-google/ [https://perma.cc/U4VN-26KQ].

[241] *See* Ian Bogost, *Go Tweak Yourself, Facebook*, ATLANTIC (Apr. 28, 2016), https://www.theatlantic.com/technology/archive/2016/04/go-tweak-yourself-facebook/480258 [https://perma.cc/7G5F-FAU2].

[242] *See* Jeffrey Gottfried & Elisa Shearer, *News Use Across Social Media Platforms 2016*, PEW RES. CTR. (May 26, 2016), http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/ [https://perma.cc/73H2-3MUJ].

[243] *Id.*

[244] *Id.*

[245] 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/8V2H-Q27S].

[246] 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, 542 (2011).

filtering out incorrect information."[247] The algorithmic audience he described is neither deliberative nor agonistic, but can be quantified and visualized[248] based on algorithms that take into account inputs like search terms, Internet traffic patterns, the ad market, keyword rates, and the competition.[249] Viewed this way, it 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. Anderson's point is, given that journalistic techniques are practices with a deep claim on democratic life, we need to conduct a broader inquiry into the "sociology and politics of algorithms."[250]

Journalism ethics and standards have developed precisely for this reason. The preamble to the Society for Professional Journalists' Code of Ethics begins with the premise that "public enlightenment is the forerunner of justice and the foundation of democracy."[251] While journalistic ethics are directed to journalists, and are mostly relevant to the content of their writings, new studies show that, today, "press ethics are intertwined with platform design ethics, and press freedom is shared with software designers."[252] As is the case with social media platforms, the design choices made by online news platforms form our experience of everyday news reading. Some have argued that designers of these apps 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."[253]

Even further, robo-journalism, or automated journalism, is now a reality.[254] The *Associated Press*,[255] *Forbes*, and the *Los Angeles*

---

247 *Id.*

248 *Id.*

249 *See id.*

250 *Id.* at 529.

251 *SPJ Code of Ethics*, SOC'Y PROF. JOURNALISTS (Sept. 6, 2014), http://www.spj.org/ethicscode.asp [https://perma.cc/J35C-KPZ9].

252 Ananny & Crawford, *supra* note 156.

253 *Id.*

254 *See* Dorrier, *supra* note 162; Joe Pinsker, *Algorithm-Generated Articles Don't Foretell the End of Journalism*, 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/Z7UC-SVHC].

Times all use some kind of platform that analyzes data and creates news reports.[256] While the technology is still at an early stage, and limited to routine stories for repetitive topics (such as sports and weather, for instance), it seems to be here to stay. While proponents see a big potential upside for journalists, who will be "free to do more reporting and less data processing" when the robots do all the drudge work, the algorithms immediately raise concerns about transparency and accountability, and potential implications for society and democracy.[257]

In its *Guide to Automated Journalism* published in January 2016, the Tow Center for Digital Journalism observed that little is known about news consumers' demand for algorithmic transparency and the extent to which they want to understand how such algorithms work.[258] But what seems unquestionable is that "automated journalism will increase the amount of available news, which will further increase people's burden to find content that is most relevant to them."[259] The report noted 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.[260] In conclusion, the report called for further research on the potential effects of personalization, the extent to which algorithms can be trusted as a mechanism for providing checks and balances, identifying important issues, and establishing a common agenda for the democratic process of public opinion formation, and the implications for democracy "if algorithms are to take over journalism's role as a watchdog for government."[261]

---

[255] *See* Erin Madigan White, *Automated Earnings Stories Multiply*, ASSOCIATED PRESS BLOG (Jan. 29, 2015), https://blog.ap.org/announcements/automated-earnings-stories-multiply [https://perma.cc/F38Y-BZBQ] (announcing that, in January 2015, the Associated Press was automatically generating more than about 3,000 U.S. corporate earnings each quarter).

[256] *See* Podolny, *supra* note 162.

[257] *Id.*

[258] ANDREAS GRAEFE, TOW CTR. FOR DIG. JOURNALISM, GUIDE TO AUTOMATED JOURNALISM (2016), http://towcenter.org/research/guide-to-automated-journalism/ [https://perma.cc/WMB9-KPTF].

[259] *Id.*

[260] *Id.*

[261] *Id.*

3. Culture

The same tensions apply not just to news but culture in general. In some ways there have always been cultural gatekeepers—in the form of music reviewers and record labels, art critics, restaurant critics—and we have always had to assess their motives. But in some other ways, what Big Data is undertaking is far more than just promotion and pop culture in a digital context. Technology critic Evgeny Morozov illustrated this "memefication" in the music industry[262] using the example of Music Xray, a new tool that allows musicians to upload their songs and have them analyzed for hit potential. The analysis is, by definition, based on previous hits. The result is that the algorithm essentially removes individuality from the creative process and determines what is being created, not simply what is being promoted.[263] Trends and movements have always existed in the music industry, making it harder to create new ones. It is not far-fetched to imagine record labels using such algorithms to filter the musicians they consider taking on, much like employers already do to screen resumes of job applicants.[264] Outliers are then even harder to discover, and the same content keeps producing itself over and over again.

C. Consequences for the Individual and Society

Going back to the big picture, when we add up individual loops or filter bubbles with the kind of trending or popular non-personalized content that is pushed on us as a result of the meme culture, we observe that the process of self-discovery is becoming increasingly controlled and directed toward the mainstream. An individual's boundary management process between the self and the social sphere is becoming less meaningful, and the breathing room that nurtures autonomous individuals is undermined.[265] Being an independent critical thinker may still be a personality trait of certain individuals, but the stimuli such individuals need to develop

---

[262] *See* MOROZOV, *supra* note 122, at 159.

[263] *See generally* Christopher Steiner, *Can Creativity Be Automated?*, MIT TECH. REV. (July 27, 2012), http://www.technologyreview.com/news/428437/can-creativity-be-automated/ [https://perma.cc/7AE4-HRAX].

[264] *See* Barocas & Selbst, *supra* note 112, at 682.

[265] *See* Cohen, *supra* note 61, at 1930–31.

their thinking are becoming increasingly difficult to discover. Referencing Julie Cohen's work on autonomy again, in such conditions, choices start inclining toward the bland and mainstream and the result is a "subtle yet fundamental shift in the content of our character, a blunting and blurring of rough edges and sharp lines . . . threatening not only to chill the expression of eccentric individuality, but also, gradually, to dampen the force of our aspirations to it."[266] In a perfectly controlled social environment, dissent is made not just impossible, but unthinkable,[267] not because of the chilling effects described in Part II, but due to lack of stimuli resulting in intellectual laziness.

Similarly, the same conditions are required for innovation in markets, as independent thought, inventiveness, and entrepreneurship cannot be stimulated inside controlled patterns.[268] More broadly, individual autonomy is a requirement for a truly democratic society.[269] Without it, democratic participation and responsible citizenship are severely undermined.[270]

Whether the issues highlighted are the same old problems, old problems with new applications, or completely new problems, the bottom line well captured by Professor Gillespie is where we need to put our focus: "We are now always navigating information and culture by way of these mechanisms, and every mechanism has a built in notion of what [it is] trying to accomplish."[271] For Gillespie, that is the part we need to unpack, namely the "assumptions that tool makes about what it should look for, what it is we seek, and [what is] important about that form of culture (whether [it is] journalism or music or whatever)."[272] Further, we need to understand the nature and complexity of the new ecosystem involved when it comes to navigating information, news, and culture.

---

266 Cohen, *supra* note 27, at 1426.

267 *See* MOROZOV, *supra* note 122, at xii.

268 Cohen, *supra* note 27, at 1427.

269 *See supra* Part I.

270 *See* Cohen, *supra* note 27, at 1426 ("Development of the capacity for autonomous choice is an indispensable condition for reasoned participation in the governance of the community and its constituent institutions—political, economic, and social. The cornerstone of a democratic society is informed and deliberate self-governance.").

271 *Id.*

272 O'Donovan, *supra* note 235.

Google has claimed to be a "virtual mirror of the world at all times,"[273] simply reflecting the state of society. The problem with that approach is that it ignores the role this mirror has in shaping the very society it claims to reflect. Media has always played a certain role in shaping society, but the difference is that people understood and expected the process of curation and could choose accordingly (people could choose to read liberal or conservative newspapers, understanding the filter under which they were reading). In some ways, what counts as relevant is just as vague as what counts as newsworthy,[274] but we do not think of Google Search as "media"; we perceive it more as a very efficient library index, and as such, we do not yet have the instinct to question the almost invisible assumptions it makes. Google prides itself on being automated, but every decision that goes into the algorithm has political questions at its heart,[275] and embodies the decisions and beliefs of Google employees who *prefer* that Google returns results that the users *believe* to be useful.[276] After all, its reputation of satisfying the users' desires and needs is how Google maintains and increases search usage, and by extension, how it sustains and increases revenues as a for-profit business.[277]

Gillespie rightly observed that there is an important tension emerging between what we expect these algorithms to be and what they in fact are, and claimed that not only must we recognize that these algorithms are not neutral, but we must also understand what it means that we are coming to rely on these algorithmic tools as our means of navigating the "huge corpuses of data that we must."[278] The truth is that we want them to be neutral, reliable, and the effective ways in which we learn what is most important, but we cannot comprehend the complexity required to create algorithms that seem to effortlessly identify what is important, without being swamped by the mundane or the irrelevant.[279]

---

273 *See* Siegler, *supra* note 145.
274 *See* O'Donovan, *supra* note 235.
275 *See generally* Grimmelmann, *supra* note 121.
276 *Id.* at 944.
277 Introna & Nissenbaum, *supra* note 141, at 176.
278 Gillespie, *supra* note 226.
279 *See id.*

In other words, we must develop the instincts to question the assumptions and values these algorithms embody and we must start undertaking this questioning. What Gillespie found troubling is that we do not have a sufficient vocabulary[280] for assessing the algorithmic intervention of such tools, nor do we have a language for the unexpected associations algorithms make beyond the intention (or even comprehension) of their designers.[281] The latter point is well illustrated in the context of discrimination, where existing language of disparate impact doctrine falls short for even describing some types of algorithmic discrimination.[282] On a much broader scale, to preserve meaningful individual autonomy, we need a clear sense of how to talk about the politics of such algorithms.[283] The questions are not just technical, but rather "sociotechnical" and have a claim to what sorts of citizens we become.[284]

IV. INFLUENCING ACTION: PERSUASIVE TECHNOLOGY

The idea that technology can influence human behavior is not new to Big Data,[285] but what *is* new is that by combining insights from psychology, neuroscience, and behavioral economics with new digital technologies and social media, companies have now moved beyond simply measuring customer behavior to creating products that are designed with the specific goal of forming new habits.[286] Examples range from determining the content and timing

---

[280] *See* Tarleton Gillespie, *Facebook, Google, and the Surprising Intricate Curation of What's Online*, SCRUTINY (Apr. 21, 2011), http://tarletongillespie.org/scrutiny/?p=121 [https://perma.cc/M35L-SYP8].

[281] *See* Gillespie, *supra* note 226.

[282] *See generally* Barocas & Selbst, *supra* note 112.

[283] *See* Gillespie, *supra* note 226.

[284] Julie E. Cohen, *Configuring the Networked Citizen*, *in* IMAGINING NEW LEGALITIES: PRIVACY AND ITS POSSIBILITIES IN THE 21ST CENTURY 129, 130 (Austin Sarat et al. eds., 2012).

[285] *See What Is Captology?*, STAN. PERSUASIVE TECH. LAB, http://captology.stanford.edu/about/what-is-captology.html [https://perma.cc/N8U9-HT47] (last visited Mar. 3, 2017). Stanford researcher B.J. Fogg devised the term "persuasive technology." *Id. See generally* STAN. PERSUASIVE TECH. LAB, http://captology.stanford.edu [https://perma.cc/39HX-VDVA] (last visited Mar. 3, 2017).

[286] *See generally Path of Persuasion*, MIT TECH. REV. 2 (May/June 2015), http://ilp.mit.edu/media/webpublications/pub/literature/tr-breports/15-08-Marketing-Influence-Persuasion.pdf [https://perma.cc/JXZ2-W8M7].

of ads on the trivial side of the spectrum, to digital games designed to keep players "hooked," to health apps prompting users to act in a certain way, to adjusting political campaign tactics.[287]

A. A/B Testing and Applications

A simple version of such techniques is A/B testing (sometimes called split testing), which entails comparing two versions of a web page to determine which one performs better.[288] One can compare two web pages by showing the two versions—an A version (the control) and a B version (the variation)—to similar visitors at the same time and measuring the effect each version has.[289] Performance is measured in terms of "conversion rate," which can be based on metrics such as sign-ups, downloads, purchases, donations, leads, registrations, user-generated content, or whatever each company's respective goals are.[290] Taken one step further, such split testing tools allow for variations of a web page to be targeted to specific groups of visitors, delivering a more tailored and personalized experience.[291] Start-ups like Optimizely, a website optimization platform that is famous for being part of the Obama campaigns, recommend constantly testing and optimizing web pages, as it provides teams with valuable insight about their visitors.[292]

Such intensive testing is now reshaping what the Internet looks like, but concerns about this can vary significantly depending on the context. For example, the fact that Google at some point tested forty-one shades of blue to see which one performs better[293] does not seem to raise any concerns. BuzzFeed, a news website that uses testing and optimization as its central principle, increases page

---

[287] *Id.* at 6.

[288] *See A/B Testing*, OPTIMIZELY, https://www.optimizely.com/ab-testing [https://perma.cc/8YZA-DZ4Q] (last visited Mar. 3, 2017).

[289] *See id.*

[290] *Id.*

[291] *Id.*

[292] *Id.*

[293] *See* Antonio Regalado, *Seeking Edge, Websites Turn to Experiments*, MIT TECH. REV. (Jan. 22, 2014), https://www.technologyreview.com/s/523671/seeking-edge-websites-turn-to-experiments/?set=523641 [https://perma.cc/S84X-KESX].

views of its "listicles" with such techniques.[294] As with algorithmic journalism described in Part III, a concern here might be that we end up with lower quality content, but it probably will not make a very convincing privacy argument against the use of such tools.

However, such methods are now not only used by start-ups, but have an increasing popularity among traditional influencers, like political campaigns. In that context, the Obama reelection campaign in 2012, which broke records for online fund-raising, is a use case that can illustrate the concerns.[295] The campaign used A/B testing to weigh every change to its fund-raising web page, and, at some point, discovered that adding a personal message from the president—"Stand with me, work with me . . ."—led to a fourteen percent increase in visitors who made an online donation to the campaign.[296] While such techniques may start feeling uncomfortable, the privacy harm is still difficult to articulate because the person affected is not the same person as the one whose information is used. The campaign used what it learns from the control group (a small number of people that are part of the test) and then applied its findings to the rest of the population whose actions might be affected.[297]

When combined with voter data, however, A/B testing enables campaigns to engage in techniques like micro-targeting, where personalized messages are targeted to individual voters.[298] Using sophisticated algorithms and modeling techniques, campaigns can infer voters' preferences, intentions, and beliefs, link personal characteristics with political beliefs, and specifically target undecided voters.[299] Robo-journalism enables a new wave of hyper-targeting;[300] tools like Facebook's new "endorsement" feature

---

294 *See* Lukas I. Alpert, *Buzzfeed Nails the 'Listicle'; What Happens Next?*, WALL ST. J. (Jan. 29, 2015, 1:38 PM), https://www.wsj.com/articles/buzzfeed-nails-the-listicle-what-happens-next-1422556723 [https://perma.cc/U5Q9-U3Q3].

295 *See* Rubinstein, *supra* note 214.

296 *Id.*

297 *See id.*

298 *Id.*

299 *See id.*

300 *See* Jonathan Holmes, *AI Is Already Making Inroads Into Journalism but Could It Win a Pulitzer?*, GUARDIAN (Apr. 3, 2016, 1:13 PM), https://www.theguardian.com/media/

make voter data easier to obtain and more accurate by turning what was previously an algorithmic inference to information provided by users themselves.[301] As the technology advances, every election cycle will come with new challenges and will force us to rethink the legality of campaigning practices. Such techniques affect the core of the democratic process, compromising values that are necessary preconditions for democratic life, such as political privacy.[302] To reframe the above discussion in the context of autonomy, we observe that these techniques interfere both with the exploration stage, by causing chilling effects[303] and by controlling the information available to each individual, and with the action/decision-making stage by steering individuals to make a campaign donation or decide to vote a certain way.

Finally, if we turn to the Facebook social contagion experiment, an application of A/B testing that has probably caused the most criticism, outrage, and debate, the autonomy concerns start becoming even more severe.[304] For a week in January 2012, the feeds of about 700,000 Facebook users were manipulated to determine how users' emotional states change depending on the nature of the

---

2016/apr/03/artificla-intelligence-robot-reporter-pulitzer-prize [https://perma.cc/4NGW-L2DQ].

[301] *See* Kate Conger, *You Can Endorse Your Preferred Presidential Candidate on Facebook Now*, TECHCRUNCH (Oct. 18, 2016), https://techcrunch.com/2016/10/18/facebook-presidential-endorsements [https://perma.cc/XR9X-NVR7]; Sofia Grafanaki, *Facebook Wants You to Get Even More Political*, ILI STUDENT BLOG (Oct. 26, 2016, 12:21 PM), http://blogs.law.nyu.edu/privacyresearchgroup/2016/10/facebook-wants-you-to-get-even-more-political/ [https://perma.cc/2TFE-BUXW]; *How Do I Endorse a Political Candidate on Facebook?*, FACEBOOK, https://www.facebook.com/help/1289003767810596 [https://perma.cc/WL64-KUGC] (last visited Mar. 3, 2017); Jeremy B. Merrill, *Liberal, Moderate or Conservative? See How Facebook Labels You*, N.Y. TIMES (Aug. 23, 2016), https://www.nytimes.com/2016/08/24/us/politics/facebook-ads-politics.html?_r=3 [https://perma.cc/47E2-XGRU].

[302] *See* Cohen, *supra* note 27, at 1426.

[303] *See id.*

[304] *See* Kate Crawford, *The Test We Can—and Should—Run on Facebook*, ATLANTIC (July 2, 2014), http://www.theatlantic.com/technology/archive/2014/07/the-test-we-canand-shouldrun-on-facebook/373819/ [https://perma.cc/QK5B-FHTA]; *see also* Robinson Meyer, *Everything We Know About Facebook's Secret Mood Manipulation Experiment*, ATLANTIC (June 28, 2014), http://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/ [https://perma.cc/M46J-6FR4].

posts they see.[305] The experiment concluded that "emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness," and "emotional contagion occurs without direct interaction between people (exposure to a friend expressing an emotion is sufficient), and in the complete absence of nonverbal cues."[306] Kate Crawford, citing sociology scholarship, convincingly argued that, in social experiments, there is a fine line that the social scientists should not cross; if they do, they are tampering with human autonomy, and ultimately exercising power and deception.[307] She quoted sociologist Edward Shills, who argued that "manipulative experimentation is not a relation between equals; it is a relationship in which power is exercised."[308]

More recently, Google's tech incubator Jigsaw developed a plan that promised to disrupt ISIS online recruiting efforts through targeted advertising.[309] The "Redirect Method" is described as a way to get inside the heads of potential terrorists before they are actually recruited, and change their intentions.[310] It works the same way as any targeted advertising campaign would. The first step is to identifying keywords and patterns of online activity that indicate a person is susceptible to or on a path toward extremism (presumably by analyzing historical data available on known terrorists).[311] Step two is to serve individuals who either search for these keywords or present similar patterns with ads (search, display, and video) that would undermine or "undo ISIS's brainwashing."[312] What seems to be a key aspect of the Redirect Method is its subtle-

---

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

[306] *See id.*

[307] *See* Crawford, *supra* note 304.

[308] *Id. See generally* EDWARD A. SHILS, THE SELECTED PAPERS OF EDWARD SHILS, VOLUME 3: THE CALLING OF SOCIOLOGY AND OTHER ESSAYS ON THE PURSUIT OF LEARNING (1980).

[309] Andy Greenberg, *Google's Clever Plan to Stop Aspiring ISIS Recruits*, WIRED (Sept. 7, 2016, 7:00 AM), https://www.wired.com/2016/09/googles-clever-plan-stop-aspiring-isis-recruits/ [https://perma.cc/E68R-7978].

[310] *Id.*

[311] *See id.*

[312] *Id.*

ty, meaning that the ads do not come across as anti-ISIS at first sight; if they did the individuals in question would not click or engage with the content. So, the ads are designed to draw people in delicately, and then subtly undermine the propaganda.[313]

Results seem to show that the program is effective—at least in that Jigsaw ads had much higher click-through rates than typical ad campaigns.[314] While there is no doubt that the cause is noble and such use is desirable, the program illustrates just how powerful Google's targeted advertising tools really are. In the context of privacy versus security, most would probably view such techniques as less intrusive than the NSA's bulk surveillance methods, for example. But could they not cause the same type of chilling effect and self-censorship? More importantly, how would we feel if the cause was not to stop terrorism, but to stop a political candidate that some deem dangerous? Would that undermine the democratic process? If, however, all that happens is that we end up buying that pair of shoes, then it feels like plain old marketing. The minute we move away from the clear cases and begin using data and analytics to get inside the minds of people and change their intentions, it starts to feel problematic because it is difficult to draw a line between acceptable and not acceptable uses. Google and Jigsaw set up the Redirect Method as a pilot program that can scale, just like any ad campaign.[315] In other words, for it to have broad reach, organizations with significant funding must pay for the ads. In fact, *any* organization with funding can use these techniques to "direct" individuals to *any* cause. This brings us back to old themes around information, power, and money, but the balance is even more skewed, giving money even more power.

B. Habit-Forming Techniques and the "Hook"

Moving beyond A/B testing and targeted marketing, companies have taken their efforts a step further, using insights from psychological and neurological research on how habits are formed to

---

[313] Jillian D'Onfro, *The Subtle Way Google Plans to Use Its Greatest Skill to Combat ISIS*, BUS. INSIDER (Sept. 11, 2016, 12:00 PM), http://www.businessinsider.com/jigsaw-redirect-method-to-stop-isis-recruits-2016-9 [https://perma.cc/6V6S-HWG8].

[314] *Id.*

[315] *Id.*

change the way consumers behave. A well-known example is the Target pregnancy case,[316] where the Target marketing team tried to change shoppers' habits by influencing them at vulnerable moments, such as the second trimester of a pregnancy.[317] A recent report from the *MIT Technology Review* on the "Path of Persuasion," provided a detailed account of these techniques, exploring "[h]ow technologies from smartphones to social media are used to influence our tastes, behavior, and even habits."[318]

The report pointed to Nir Eyal, a technology entrepreneur who writes extensively on the intersection of technology, psychology, and business, and has been very influential in the development of habit-forming technology.[319] Based on existing research, he advocated for a technique he called "the hook,"[320] a four-step process starting with (1) a *trigger*, which prompts one to (2) *take action* that leads to (3) a *reward*, which then causes the user to inject a personal stake by (4) *making an investment*, thereby closing the loop by "loading the next trigger."[321] In a simple example used to illustrate the process, a user receives a Facebook notification indicating they were tagged in a photo (trigger), prompting them to log on to Facebook (action) to view the photo (reward) and make a comment (investment), a reply to which puts them right back in the beginning of the loop.[322] According to the research, such techniques speak to the brain's reward center by triggering a release of dopamine and creating new neural pathways and reward circuits.[323] What is especially troubling is that this reward circuitry of the brain is also the mechanism behind addictions[324] and companies are now specifical-

---

316 *See* Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAG. (Feb. 16, 2012), http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0 [https://perma.cc/W8NR-RM27]; *see also* CHARLES DUHIGG, THE POWER OF HABIT: WHY WE DO WHAT WE DO IN LIFE AND BUSINESS 182–83 (Random House 2012).

317 Duhigg, *supra* note 316.

318 *See Path of Persuasion*, *supra* note 286, at 1.

319 *See generally About Nir Eyal & NirAndFar.com*, NIR & FAR, http://www.nirandfar.com/about [https://perma.cc/J3LD-PV2D] (last visited Mar. 3, 2017).

320 *See Path of Persuasion*, *supra* note 286, at 6–9.

321 *Id.*

322 *See id.* at 7.

323 *See id.* at 6; *see also* DUHIGG, *supra* note 316, at 19.

324 *See* AMERICAN SOCIETY OF ADDICTION MEDICINE, *Definition of Addiction*, http://www.asam.org/for-the-public/definition-of-addiction [https://perma.cc/R9AB-3BNF]

ly designing products and services that generate compulsive behavior.[325]

In some contexts, the use of such techniques may not seem so alarming. Casinos have used such techniques for a long time, with slot machines specifically designed based on these ideas.[326] Now, companies like Expedia are trying to get users to return to their website daily by developing tools like the Scratchpad, and game designers are now talking about forming a "compulsion loop."[327] The effect can even be positive when used in the health context, with wearable technology companies like Jawbone trying to understand what gets people to act, and encourage healthy habits like more exercise and better sleep patterns.[328] That said, there are no clear limits for such persuasion techniques, and when combined with the increasing amount of information about us that companies know, can track, or can infer, they can start looking more like manipulation than persuasion.

Persuasion profiling essentially combines the algorithms that create the loops discussed in Part III with testing and persuasion technologies. A simple version of this may be price discrimination, when companies can perform extensive testing to understand different consumers' willingness to pay, and then adjust prices on an individual basis accordingly.[329] So far, the topic has not been met

---

("Addiction is a primary, chronic disease of brain reward, motivation, memory and related circuitry.").

[325] *See Path of Persuasion*, *supra* note 286, at 6. The author elaborated further on the ideas behind designing such products. *Id.* at 6–7. The hook technique is further refined based on the work of Stanford behavioral theorist B. J. Fogg, according to whom a behavior happens when a trigger coincides with both motivation and ability, but only when they are in the right proportion. *Id.* at 7. The author explained:

> If a trigger consistently fails to initiate the desired action, the theory goes, habit designers should aim to enhance the user's ability. Motivation is hard to influence, because you [cannot] make people do what they [do not] want to do. Ability is more malleable: simply make the behavior easier to execute.

Id.; *see also* NIR EYAL, HOOKED: HOW TO BUILD HABIT-FORMING PRODUCTS 71 (CreateSpace 2013).

[326] *See Path of Persuasion*, *supra* note 286, at 6.

[327] *Id.* at 8.

[328] *See id.* at 12.

[329] *See generally* Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995 (2014).

with great concern and economic theory can provide us with market justifications for it,[330] but when we move from an abstract understanding of the practices to specific use cases, there is a point where we start feeling our autonomy severely compromised. Eli Pariser illustrated that, when combined with methods of sentiment analysis that allow the observer to understand what mood an individual is in, the outcome can move away from simply tailoring content to someone to take advantage of his psychology.[331] Not only can companies infer whether someone is happy or sad based on his communications, but they can also distinguish sober messages from drunk ones, for example, by analyzing the amount of typos in the communication.[332] Implications can vary from simply having content tailored to one's mood, which (as illustrated by the Facebook emotional contagion experiment) can keep the person in a "feeling sad" loop if the filter shows more negative content, to triggering compulsive purchases if they can infer that a particular individual is more susceptible at certain times,[333] or during certain states, such as feeling sad or being tipsy. Like Crawford, Pariser also discussed power and deception, pointing out the inherent information asymmetries in the process as well as the fact that, unlike other forms of profiling, persuasion profiling is handicapped when it is revealed.[334] He illustrated this by giving an example of an automated coach saying, "[You are] doing a great job! [I am] telling you that because you respond well to encouragement!," which surely cannot have the same effect as a plain, "You can do it!"[335] The effects of transparency to the effectiveness of the technique may vary depending

---

330 *See* Jason Furman & Tim Simcoe, *The Economics of Big Data and Differential Pricing*, WHITE HOUSE BLOG (Feb. 6, 2015, 3:58 PM), https://obamawhitehouse.archives.gov/blog/2015/02/06/economics-big-data-and-differential-pricing [https://perma.cc/U33S-6Y7J].

331 *See* PARISER, *supra* note 122, at 121.

332 *Id.*

333 *See* Kramer, *supra* note 306, at 8788; *see also Ovulation Hormones Make Women 'Choose Clingy Clothes,'* BBC NEWS (Aug. 5, 2010), http://www.bbc.com/news/health-10878750 [https://perma.cc/FDA8-7YMK] (discussing discovery by researchers at the University of Minnesota that women who are ovulating respond better to pitches for clingy clothes and suggested that marketers "strategically time" their online solicitations).

334 *See* PARISER, *supra* note 122, at 123.

335 *Id.*

on the context, but what is apparent is a lack of incentives for transparency from the side of companies, especially given the investment needed to develop these methods, both financially and temporally.

Consequently, if we reframe these observations in terms of individual autonomy, we see that independent action based on an independent exploration process (i.e., independent and rational decision-making) can be completely circumvented with the use of these technologies. We can again have the appearance of free choice, but a choice based on psychological maneuvering (if not manipulation) is far from meaningful.

CONCLUSION

What this Article attempted to illustrate is that from chilling effects, to self-reinforcing loops (or filter bubbles) on individuals and society, to techniques for persuasion profiling, Big Data tools come with the danger of slowly and gradually nudging individuals to a preset scheme, and inhibiting their individuality. Such concerns are not new to Big Data; they were expressed long ago when computerization and automated processing started being used for individuals' data. That said, the exponentially growing complexity of the tools has introduced challenges that existing legal frameworks such as control, choice and consent, transparency, and information accuracy cannot address; they fail to protect individuals from harm and they are not adequate safeguards to meaningful individual autonomy. While solutions are beyond the scope of this Article, we end with some concluding observations.

At its infancy, the Internet was seen as something like an "ideal" or a "platform for social justice," promising to be a democratizing force by giving voice and access to diverse socioeconomic and cultural groups, empowering the traditionally disempowered, and enabling people to communicate and associate in ways they had never done before.[336] As Harvard Law professor and renowned author Lawrence Lessig put it, cyberspace "promised a kind of so-

---

[336] Introna & Nissenbaum, *supra* note 141, at 181; *see also* LAWRENCE LESSIG, CODE: AND OTHER LAWS OF CYBERSPACE at 4 (1999).

ciety that real space could never allow—freedom without anarchy, control without government, consensus without power," and it was thought, by its very nature, to be "free."[337] Critics often label this conception as cyber-utopianism.[338]

For Lessig, cyberspace has no nature, meaning "no particular architecture that cannot be changed. Its architecture is a function of its design . . . its code."[339] As such, liberty in cyberspace will not just emerge, but will come from foundations of a certain kind. Just as society was set upon a certain constitution in order to protect fundamental values and structure and constrain power, the same kind of foundations are required in cyberspace if it is to fulfill the promise of freedom.

Drawing from several scholars, this Article illustrated how embedded politics, beliefs, and values are in our devices and technologies. Political theorist and leading academic on the politics of technology Langdon Winner, writing in 1980, argued that artifacts have politics,[340] Fordham Law professor and Internet law scholar Joel Reidenberg in 1997 talked about *Lex Informatica*,[341] and Lessig in 1999 argued that "[c]ode is law."[342] While these texts have been widely quoted, it seems that, as a society, we are only starting to see how this "code" affects us, perhaps because a fraction of the harms that previously sounded too abstract have started materializing. But we still lack the vocabulary to properly talk about it[343] and more fundamentally we lack a clear set of ethical norms around which to formulate the discussion.[344]

---

337 LESSIG, *supra* note 336, at 4–5.

338 *See* EVGENY MOROZOV, THE NET DELUSION: THE DARK SIDE OF INTERNET FREEDOM xiii (2012).

339 Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501, 501 (1999).

340 *See* Langdon Winner, *Do Artifacts Have Politics?*, 109 DAEDALUS 121, 121 (1980).

341 *See* Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553, 553 (1998).

342 LESSIG, *supra* note 336, at 6.

343 *See* Gillespie, *supra* note 226.

344 *See* Paul Ohm, *Changing the Rules: General Principles for Data Use and Analysis*, *in* PRIVACY, BIG DATA, AND THE PUBLIC GOOD FRAMEWORKS FOR ENGAGEMENT, *supra* note 6, at 96, 108.

At the heart of the matter, lies individual autonomy, without which, democracy cannot exist, neither in real space nor in cyberspace. For Lessig, we can *code* cyberspace to either protect values that we believe are fundamental, or to allow those values to disappear.[345] This is not a technophobic critique, and nothing in this Article was meant as such. Rather, drawing from several scholars and critics, this Article attempted to illustrate the extent to which politics, beliefs, and values are embedded in our devices and technologies, and how the consequences can range from trivial to very severe depending on the context.[346]

Fortunately, we do not yet live in a singularity with no comprehension or control of the progress of machines. The Internet is still "a human invention and can be altered by humans,"[347] or put differently, "code can change, either because it evolves in a different way, or because government or business pushes it to evolve in a particular way."[348] Entrepreneurs have recently claimed that evolution in code is already under way, brought on by computational designers (or "code artists") who are advancing a new computer language described as object-based or "generative" code, which may able to liberate us from "systems that manipulate and control the data flows to return the most banal, and insidious, behavioral insights."[349] Even if true, any new code will have its own embedded values, and if we want to preserve the ideal of the Internet as a public good,[350] the complexities involved make it almost imperative that both the architects (coders) and the regulators work together to develop the ethical norms that will govern.

---

345 LESSIG, *supra* note 336, at 6.

346 *See generally* HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE (2009).

347 PASQUALE, *supra* note 120, at 196. Pasquale based this conclusion on the observation that technologies in general are "just as much a product of social, market, and political forces as they are the outgrowth of scientific advance," because they are "intimately embedded with social practices that rely on human judgment." *Id.* at 197.

348 Lessig, *supra* note 339, at 506.

349 Stephen Marshall, *Engineering God Mode*, LINKEDIN (Apr. 30, 2015), https://www.linkedin.com/pulse/engineering-god-mode-stephen-marshall [https://perma.cc/YJL8-RFB4].

350 *See* Introna & Nissenbaum, *supra* note 141, at 181–82.