These are proofs for an article to be published in 2015.

Original Research Article

Big Data & Society
July–December 2015: 1–15
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2053951715608876
bds.sagepub.com
SAGE

Big Data and The Phantom Public: Walter Lippmann and the fallacy of data privacy self-management

Jonathan Obar

Abstract

In 1927, Walter Lippmann published *The Phantom Public*, denouncing what he refers to as the 'mystical fallacy of democracy.' Decrying romantic democratic models that privilege self-governance, he writes: "I have not happened to meet anybody, from a President of the United States to a professor of political science, who came anywhere near to embodying the accepted ideal of the sovereign and omnicompetent citizen". Almost 90-years later, Lippmann's pragmatism is as relevant as ever, and should be applied in new contexts where similar self-governance concerns persist. This paper does just that, repurposing Lippmann's argument in the context of the ongoing debate over the role of the digital citizen in Big Data management. It is argued that recent proposals by the Federal Trade Commission, the White House and the US Congress, championing notice and choice privacy policy, perpetuate a self-governance fallacy comparable to Lippmann's, referred to here as the *fallacy of data privacy self-management*. Even if the digital citizen had the faculties *and* the system for data privacy self-management, the digital citizen has little time for data governance. What we desire is the freedom to pursue the ends of digital production, without being inhibited by the means. The average digital citizen wants privacy, and safety, but cannot complete all that is required for its protection. The Big Data self-governance challenge should be seen as an extension of a longstanding self-governance debate, exemplified by the works of John Dewey and Walter Lippmann. Both framed their arguments in opposition to a system struggling to find an autonomous and efficacious role for the citizen. If it is true that the fallacy of democracy is similar to the fallacy of data privacy self-management, then perhaps the pragmatic solution is *representative data management*; a combination of non/for-profit digital dossier management via infomediaries that can ensure the protection of personal data, while freeing individuals from what Lippmann referred to as an 'unattainable ideal.'

Keywords

Big Data, privacy, data management, information policy, data privacy self-management

The digital citizen today maintains a perpetual information illiteracy—an intellectual detachment from the rapidly expanding universe of Big Data. The digital citizen knows they are somehow affected by what is going on. Internet evolution continually, terms of service statements regularly, and data privacy mentions occasionally, serve as reminders that they are being swept along by great drifts of circumstance.

Yet the Internet's data-driven affairs are in no convincing way the affairs of the digital citizen. Big Data's operations are for the most part invisible, managed at distant centers from behind the scenes, by unnamed powers. As a private person, the digital citizen does not know for certain what is going on, or who is doing it, or where they are being carried. No newspaper reports their environment so that they can grasp it; no school has taught them how to imagine it; their ideals, often, do not fit with it. Digital citizens live in a world which they cannot see, do not understand and are unable to direct. In the cold light of experience the

Canada

Corresponding author:
Jonathan Obar, Canada.
Email: obar@msu.edu

Creative Commons NonCommercial-NoDerivs CC-BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License (http://www.creativecommons.org/licenses/by-nc-nd/3.0/) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

> digital citizen knows that data privacy self-management is a fiction.

Had *The Phantom Public* been written today, and Walter Lippmann been concerned with the nature of our digital existence, perhaps this is how his opening chapter would have read. In 1927, critiquing what he refers to as "the mystical fallacy of democracy," (p. 28) Lippmann decries academic and popular misconceptions suggesting that individuals can be self-governing in a democracy. With language now infamous for its fiery and at times offensive tone, Lippmann challenges perpetuations of democratic delusion; spoken sedatives that, to this day, are strategically prescribed by politicians aiming to assuage publics that might otherwise fear political marginalization. Almost ninety years later, his words and his pragmatism are as relevant as ever, and should be applied in new contexts where similarly romantic and impractical calls for self-governance quiet those easily sedated, concurrently wasting time, money and energy in the pursuit of what Walter Lippmann called, an "unattainable ideal." (p. 29)

This paper does just that, repurposing Lippmann's argument in an attempt to contribute to the self-governance debate taking place over the role of the digital citizen[1] in their own Big Data management. It is argued here that recent calls for data privacy self-management, or the ability for a single individual to control how their personal data is collected, used and disclosed (Solove, 2012), reveals a self-governance fallacy comparable to the fallacy described by Lippmann. What I term *the fallacy of data privacy self-management*, or the misconception that digital citizens can be self-governing in a digital universe defined by Big Data, is perpetuated by governments the world over, refusing to move beyond flawed notice and choice policy. While digital citizens suffer reputation management woes (Citron, 2009), self-disclosure misappropriation (Noamgalai.com, 2015), revenge porn (Citron and Franks, 2014), identity theft (Solove, 2002), eligibility threats from algorithms and data brokers (Pasquale, 2015), and a "swarming confusion of (other) problems" (Lippmann, 1927: 14) linked to the exponential growth of Big Data, governments champion futile 'notice' efforts in the name of privacy, engendering 'the biggest lie on the internet',[2] and data management practices ('choice' and 'access'), that limit individual data control more often than not (e.g. Parsons, 2014). In an attempt to contribute to the scholarship already highlighting the flaws in notice and choice privacy policy (e.g. Ben-Shahar and Schneider, 2011; McDonald and Cranor, 2008; Solove, 2012), this paper applies Lippmann's self-governance concerns to further demonstrate the futility of the current approach. In doing so, the intention is to strengthen the community of critique by connecting the current Big Data self-governance debate to the rich and longstanding literature addressing the role of the individual in societal governance; a debate that can be traced back at least to the ancient Greeks. By making these connections, hopefully ongoing and future privacy efforts will learn and draw from the long history of self-governance concern, and do more to champion pragmatic approaches to Big Data management that are beneficial to digital citizens.

This paper begins with a review of Lippmann's 'fallacy of democracy,' allowing for further conceptualization of the 'fallacy of data privacy self-management.' What follows is a policy analysis of recent privacy efforts by the US government that perpetuate the data privacy self-management fallacy. The analysis begins with a look at data proliferation, linked to a sweeping digitization of everyday life, and a Big Data industry that is growing as quickly as its stockpiles. The myriad of data sources and data collectors will be emphasized in an attempt to highlight the complexity and impossibility of data privacy self-management. Calls for data self-governance by the Federal Trade Commission (FTC), the White House, and the US Congress are presented next, with brief references to Lippmann's critique interspersed. Recent research supporting the view that data privacy self-management is a fallacy will be described. The discussion section summarizes the critique and briefly introduces a more pragmatic approach to the challenges identified, and one in need of further inquiry—representative data management.

Lippmann's 'Fallacy of democracy' and the 'Fallacy of data privacy self-management'

Critiquing what he views as the misconception that individuals can be self-governing in a democracy, Lippmann writes, "I think it is a false ideal. I don't mean an undesirable ideal. I mean an unattainable ideal [. . .] An ideal should express the true possibilities of its subject" (p. 29). Lippmann explains that coming to terms with the fatal flaws[3] common to models of participatory democracy reveals the fallacy. The flaws most relevant to our current inquiry, among those that Lippmann suggests limit the true possibilities of the subject, are individualistic as well as structural. Speaking to the limitations of the individual, he writes:

> "I have not happened to meet anybody, from a President of the United States to a professor of political science, who came anywhere near to embodying the accepted ideal of the sovereign and omnicompetent citizen." (Lippmann, 1927: 11)

Beyond the challenge of ubiquitous expertise in an increasingly technocratic world, Lippmann argues that

it is the impossibility of omnicompetence, expertise in all the areas of government, that is the one among the fatal flaws. He writes:

> So I have been reading some of the new standard textbooks used to teach citizenship in schools and colleges. After reading them I do not see how any one can escape the conclusion that man must have the appetite of an encyclopaedist and infinite time ahead of him. [...] He is told, in one textbook of five hundred concise, contentious pages [...] about city problems, state problems, national problems, international problems, trust problems, labor problems, transportation problems, banking problems, rural problems, agricultural problems, and so on *ad infinitum*. (pp. 13–14)

Lippmann includes himself among those that stand little chance in the face of an ever-increasing multitude of questions to be answered,

> My sympathies are with him, for I believe that he has been saddled with an impossible task and that he is asked to practice an unattainable ideal. I find it so myself for, although public business is my main interest and I give most of my time to watching it, I cannot find time to do what is expected of me in the theory of democracy; that is, to know what is going on and to have an opinion worth expressing on every question which confronts a self-governing community. (p. 10)

A second fatal flaw similarly revealed by the long list of tasks for which the self-governor would be responsible is the limitation of individual time. Lippmann argues that having the required faculties at our disposal, not to mention the will to engage with all issues would still be futile, as the lack of time available to address and answer all governance questions would leave society at a standstill. Speaking again to the civics teachers, he writes,

> [N]owhere in this well-meant book is the sovereign citizen of the future given a hint as to how, while he is earning a living, rearing children and enjoying his life, he is to keep himself informed about the progress of this swarming confusion of problems. (p. 14)

A third fatal flaw Lippmann associates with models of participatory democracy that privilege self-governance, speaks to the lack of a pragmatic structure or system that would make self-governance possible. Indeed, how big an Ecclesia is needed to house 300 million self-governors?[4] What type of interface would allow for such a crowd to raise all issues debate and resolve them? For Lippmann, this structural concern is a by-product of the lack of omnicompetence and time necessary to answer all questions. Furthermore, there simply isn't a system or interface that would allow all self-governors to answer all questions quickly enough to affect all things all the time, especially as they evolve each day. He writes,

> It never occurs to this preceptor of civic duty to provide the student with a rule by which he can know whether on Thursday it is his duty to consider subways in Brooklyn or the Manchurian Railway, nor how, if he determines on Thursday to express his sovereign will on the subway question, he is to repair those gaps in his knowledge of that question which are due to his having been preoccupied the day before in expressing his sovereign will about rural credits in Montana and the rights of Britain in the Sudan. Yet he cannot know all about everything all the time, and while he is watching one thing a thousand others undergo great changes. (p. 15)

Much has been written about Lippmann's critique of romantic preoccupations with self-governance. Relevant to the current discussion is Lippmann's supposed denigration of individual autonomy and efficacy, contributing to allegations that his position was anti-democratic (Schudson, 2008). For example, James Carey has linked Lippmann's work to a problematic shift in approaches to democratic theory that view the public as incompetent and incapable of participation in the governance process (e.g. Carey, 1987). John Dewey's writings, most notably *The Public and its Problems* (1927), championing a more inclusive participatory model have often been placed in opposition to Lippmann's. Whether or not there ever was a "Lippmann–Dewey debate" as some have claimed and others disputed (Schudson, 2008), both Lippmann and Dewey framed their arguments in opposition to what they viewed as a struggling American democracy, prescribing different, but potential solutions for the American political system (Whipple, 2005). What matters for our current purposes is to say that the differences in their assertions, and Dewey's close proximity to the romantic ideal, do not invalidate Lippmann's argument. Dewey himself understood this, as noted in his review of *The Phantom Public* (Dewey, 1925):

> While one might cite passages which, if divorced from their context, would give the impression that Mr. Lippmann was permanently "off" democracy, Mr. Lippmann's essay is in reality a statement of faith in a pruned and temperate democratic theory, and a presentation of methods by which a reasonable conception of democracy can be made to work, not absolutely, but at least better than democracy works under an exaggerated and undisciplined notion of the public and its powers. (p. 52)

Dewey appears to appreciate the attempt to move beyond an exaggerated false ideal—or as Lippmann states, "I do not mean an undesireable ideal. I mean an unattainable ideal." (Lippmann, 1927: 29). Lippmann's work attempts to identify the unattainable and pursue the pragmatic. Lippmann's work is anti-democratic only to the extent that it upsets those wedded to the warmth of a mystical democratic delusion. Considering the fatal flaws identified by Lippmann, it could be suggested that his pragmatism does more to champion citizen empowerment than does Dewey's call for direct participation, through his critique of a governance model that does little the engender practical self-governance outcomes. As Lippmann wrote:

> *An ideal should express the true possibilities of its subject* (emphasis added). When it does not it perverts the true possibilities. The ideal of the omnicompetent, sovereign citizen is, in my opinion, such a false ideal. It is unattainable. The pursuit of it is misleading. The failure to achieve it has produced the current disenchantment. (1927, p. 29)

Repurposing Lippmann and extending the self-governance debate to the Big Data context, it is argued here that each of the self-governance flaws identified (lack of omnicompetence, time and structure) are similarly revealed by the fallacy of data privacy self-management. Discussed in greater detail in the policy analysis that follows, calls for data privacy self-management, or the ability for a single individual to control how their personal data is collected, used and disclosed (Solove, 2012), highlight comparable self-governance challenges to those identified by Lippmann, and a correspondingly similar demand for pragmatic alternatives. As will be discussed in the next section, the multitude of tasks that the individual user must address in the Big Data context presents a comparable omnicompetence challenge. Beyond traditional and expanding digital divide concerns (e.g. Hargittai and Hinnant, 2008; Napoli and Obar, 2014a, 2014b; Prieger, 2013), it is unrealistic to expect ubiquitous omnicompetence in the form of understanding and continuous management of data being collected, organized, analyzed, as well as repurposed and sold by every application, commercial organization, non-commercial organization, government agency, data broker and third-party, while also expecting users to read and understand every terms of service (TOS) agreement and privacy policy. New research also suggests that data drawn from the physical layer of the Internet is relevant to the digital citizen (Clement and Obar, 2015b; Obar and Clement, 2012), adding knowledge of Internet infrastructure, management and operations to the list of tasks requiring oversight.

Though more research in the area is needed, McDonald and Cranor (2008) revealed a number of years ago that, at the time, before the explosion of social media, touch-screen phones and tablets, it would take users an average of 40 min a day to read all the privacy policies that they encounter. This alone suggests a time management concern associated with self-governance in the Big Data universe. Imagine how much additional time would be needed to manage all the tasks continually introduced by Big Data. After engaging with this "swarming confusion of problems" (Lippmann, 1927: 14) would digital citizens have time to actually use the Internet? To work? To have a family? To do anything else?

The lack of a structure or system for enabling digital citizens to manage their Big Data is also highly problematic. Beyond attempts to improve the informed consent process for privacy and TOS policies (e.g. Microsoft, 2015), which is perhaps the first of many, many steps required for data privacy self-management, it is clear that the speed of analysis as well as change within the Big Data industry is dizzying. The four V's (volume, velocity, variety and veracity) have been used consistently to draw a ring around the Big Data concept (see IBM, 2015a). Some have suggested that it is velocity, the speed at which data is being collected, analyzed and utilized that is currently the key to Big Data's appeal. Burn-Murdoch (2013) notes "it is speed, not size, that defines big data in 2013," and "it is speed, not size that is increasingly driving desire for software and hardware improvements at data-processing organisations." The reason velocity is so important is the desire for real-time and predictive results. IBM in particular is clear about its emphasis on speed, and how its products will help customers achieve their velocity-oriented needs, "the most advanced analytics software in the world doesn't do you any good if (it) takes forever to get insights. You need to have the right infrastructure in place to be able to get the most from big data and analytics in realtime" (IBM, 2015b). They add, "those that are able to acquire, analyze and act on data faster than their competitors will be the ones actually able to benefit the most from those efforts" (Reese, 2014). Speaking about one of their Big Data products, IBM Spark, the justification for speed is clear,

> As the speed of business keeps accelerating, the value of knowing what *happened* pales in comparison to the value of knowing what is *happening* right now. Imagine the power of knowing how your ad campaign is performing this very minute. Imagine the benefit of knowing where sales are spiking or tanking every day; where inventory is low or high as sales are made; and which patients are responding to care and which are not—all right now. (Howes, 2015)

In addition to the velocity issue, the speed of change in the Big Data industry is also a concern. In August of 2012, Facebook's Vice President of Engineering, Jay Parikh provided a glimpse into the industry's attitude towards Big Data's constant evolution, stating, "(in a few months) no one will care you have 100 petabytes of data in your warehouse [...] the world is getting hungrier and hungrier for data" (Constine, 2012: 4). Where does the average user fit in this whirlwind of speed and change? Speaking to his own concerns about the speed of change in the context of societal governance, Lippmann (1927) wrote:

> [...] he cannot know all about everything all the time, and while he is watching one thing a thousand others undergo great changes. Unless he can discover some rational ground for fixing his attention where it will do the most good, and in a way that suits his inherently amateurish equipment, he will be as bewildered as a puppy trying to lick three bones at once." (p. 15)

The fatal self-governance flaws of omnicompetence, time and structure saddle both the expert and the novice user with an impossible task in a digital universe defined by Big Data. Indeed, the challenge of citizen empowerment embodied by the fallacy of data privacy self-management should be understood as an extension of the self-governance challenges identified by Lippmann's fallacy of democracy. Nevertheless, recent efforts by the US government continue to pursue flawed notice and choice policies (discussed further on) that cling to romantic self-governance ideals and perpetuate the fallacy. *A plan for data privacy self-management should express the true possibilities of its subject*. Achieving pragmatic ends to empower digital citizens will not be easy, and will first require movement beyond romantic notions that remain as impossible as direct democracy within a nation of millions.

The Big Data boom

A description of the individual's relationship to the Big Data boom should be prefaced with the following: general details about Big Data and the number of potential entities involved, do not address the extent to which individuals are implicated in the practice of data collection, management, retention, sharing, etc. which remains for the most part, a mystery. The enigmatic nature of Big Data is due to a number of factors. First, in 2011, the McKinsey Institute reported that every sector in the global economy is now addressing the role of Big Data (Manyika et al., 2011). Thus the Economist's comment "*Data, Data, Everywhere*" (The economist, 2010) should be interpreted to mean not only that data is being collected everywhere, but also that a multitude of organizations are engaging with the growing Big Data industry, each presenting their own mosaic of Big Data questions. This suggests that within and across all sectors of the global economy, data is being collected, organized and retained in unique ways, contributing to a relative inability to standardize descriptions of Big Data practices for the purpose of user understanding.

Second, the lack of transparency on the part of organizations engaged with the Big Data industry further contributes to an inability to understand the extent to which the individual is implicated. Recent empirical assessments of Internet service providers (ISPs), Internet carriers, and other online intermediaries suggest that organizations are closed-lipped about the specifics of data collection, retention and management (Cardozo et al., 2014; Clement and Obar, 2014, 2015a). The growing trend in transparency reporting (e.g. recent transparency reports from Microsoft, Twitter and Comcast) provides insight into the practice of data disclosure requests from security agencies and the extent of compliance. Unfortunately, few details relevant to individual accounts are made available, again, making it difficult for digital citizens to assess their own connection to data disclosures. The flawed notice and choice policies that many of these companies do mention in their TOS and privacy policies provide users with the opportunity to request access to personal data. Beyond the impracticality of this option for common use by the general public (e.g. Ben-Shahar and Schneider, 2011; Solove, 2012), recent efforts suggest that when individuals do make the effort, they are met with a variety of obstacles (e.g. Parsons, 2014).

Concerns with transparency also apply to the growing industry of data brokers, commercial entities that collect information about individuals and then organize and package that information to sell to another party (Pasquale, 2015). A recent report by the FTC noted:

> Data brokers acquire a vast array of detailed and specific information about consumers; analyze it to make inferences about consumers, some of which may be considered sensitive; and share the information with clients in a range of industries. All of this activity takes place behind the scenes, without consumers' knowledge. (2014: vii)

While companies like Facebook and Google are in the business of interacting with the public, data brokers are not as they communicate directly with organizations interested in purchasing data. Beyond the challenge of dealing with digital dossiers at organizations we knowingly interact with, users are also faced with the difficulty of identifying dossiers managed by hidden data brokers.

The purpose of this preface was to emphasize how difficult it is for the digital citizen to develop both a detailed understanding of the complexities of Big Data, as well as a clear picture of their own digital footprint, before attempting to overcome the challenges of omnicompetence, time and structure.

Big Data by the numbers

Though Web 2.0, the mobile industry and the 'Internet of Things' have amplified the amount of information being collected, data collection practices have existed for quite some time. Census data and other government records, police records, health records, credit reports, credit card data and other information collected by financial institutions, highlight just a few of the traditional sources of data. Newer sources also exist. Schneier (2015) describes how all of our interactions with computers produce data of some kind. The proliferation of computers in the home, in public, at work and even on the person highlights the omnipresence of data collection devices. The Internet in general, and social media in particular, are major sources of data. Clickstream collection and other forms of web tracking, practices that have been common since the 1990s, are just the beginning of the data deluge. All user-generated content including text, links, photos, audio and video, 'likes,' social network lists, game results, selections, deselections, and an evolving number of other behavioral manifestations online are all being collected. The myriad possibilities from mobile phone data are also being realized. As Schneier (2015) describes, mobile phones not only track where you live, work, spend evenings and weekends, they also can assess your relationship to other mobiles in the area, suggesting the capability for determining whom you are interacting with in person as well as online.

Beyond the Global Positioning (GPS) capabilities of your phone, many mobile phones are also smartphones—a device that performs many functions commonly associated with computers. All of the applications that individuals use when engaging with a smartphone produce data. Wearable media devices, still in their infancy, like the Apple iWatch and the Fitbit also introduce new possibilities for constant tracking. Schneier (2015) describes a few more of the many computers we encounter daily including the machines used in stores to identify loyalty cards and process credit/debit purchases, computers in cars generating data about location, driving quality and quantity, as well as video surveillance devices monitoring indoor and outdoor spaces.

In 2011, a McKinsey Institute report presented an early description of what is now referred to as the 'Internet of Things', highlighting the proliferation of tracking technologies,

> More than 30 million networked sensor nodes are now present in the transportation, automotive, industrial, utilities, and retail sectors. The number of these sensors is increasing at a rate of more than 30 percent a year. (Manyika et al., 2011: 2)

Coupled with these tracking technologies are also facial and gait recognition applications that help analyze and infer from data (e.g. Acquisti et al., 2014; Schneier, 2015). Emphasis on data analysis technologies such as predictive analytics suggests that our attention should focus not only on data collection, but data aggregation, management, retention, disclosure, etc. This further elucidates that digital citizens are tasked with understanding not only a multitude of diverse challenges associated with data collection, but also a variety of aggregation and analytical possibilities.

The term being used to describe the constantly and exponentially evolving universe of information being collected is 'Big Data.' The 'iceberg' cliché applies to the term, as users are only periodically given glimpses into Big Data collection and management practices, typically hidden from public view.

For example, in May of 2012, Bamford (2012) published an article in *Wired* about a new multi-billion dollar National Security Agency (NSA) "spy center" being built in the Utah desert capable of handling yottabytes of data. A yottabyte is one septillion (10^{24}) bytes; the progression moves from gigabyte (10^{9}) to terabyte (10^{12}) to petabyte (10^{15}) to exabyte (10^{18}) to zettabyte (10^{21}) to yottabyte. In 2009, it was estimated that the entire global Internet contained 500 exabytes of data (Wray, 2009), further emphasizing Big Data's rate of growth. In August of 2012, Facebook's Vice President of Engineering, Jay Parikh, reported that Facebook was accepting 2.5 billion pieces of content (including 300 million photos) and more than 500 terabytes of data each day (Constine, 2012). Facebook's Big Data represents the collection practices of just one site. Countless other sites visited by users every day including Google, Twitter, Yahoo, Amazon, MSN, CNN.com are all amassing stockpiles of their own. Furthermore, Internet companies are hardly the only entities interested in Big Data. As noted, Manyika et al. (2011) reported that every sector in the global economy is now addressing Big Data questions.

A discussion of the Big Data industry must include a mention of data brokers. Data brokers, also referred to as data aggregators, information brokers or data vendors, collect information about individuals, then organize and package that information for the purpose of selling data to another party. Their methods of data collection are varied and controversial. Methods range from scraping public data like names, contact information and other user-generated content from

publicly accessible locations, especially on the Internet, to the acquisition of purchasing histories, credit card activities, registrations with commercial and non-commercial organizations, charitable and religious affiliations, as well as bulk-data purchases of massive databases from government (and likely) private entities (Couts, 2012; Mitchell, 2012; Wayne, 2012).

3

In summary, IBM estimates that by 2020, a total of 40 zettabytes (10^{21}) of data will have been created, an increase of 300 times from 2005 (IBM, 2015a). Most companies operating in the US have at least 100 terabytes of data stored, a number that is also increasing, and by 2016 there will be approximately 19 billion network connections across the world, 2.5 connections per individual on Earth (IBM, 2015a).

With such a wide variety of entities involved in the collection, management and trade of Big Data, data privacy self-management proposals must go beyond broad and ambiguous notice and choice possibilities. Proposals without pragmatics for addressing our lack of omnicompetence, time and structure are at best a first step, and at worst, futile.

Calls for data privacy self-management

In 1973, responding to the growth of digital data collection, the FTC released a set of Fair Information Practice Principles (FIPP), intended to guide industry practice as well as the development of law and policy. In the years that have followed, a variety of privacy laws in the United States and around the world (in Canada and in Europe in particular) were developed in accordance with the FIPPs.[5] The principles include the following: 1) transparency of data record systems, 2) the right to notice about data record systems, 3) the right to prevent data from being used without consent, 4) the right to correct or amend personal data, and 5) that data holders are responsible for the safekeeping of data and to ensure that data isn't misused (c.f. Solove, 2012). The FIPPs are now commonly referred to as notice, choice, access, security and enforcement (McDonald and Cranor, 2008). As has been previously discussed, the first three FIPPs (often characterized as 'notice and choice' policy) are proving to be problematic in the context of Big Data. The notice principle raises concerns due the difficulty achieving informed consent with privacy and TOS policies (McDonald and Cranor, 2008). The choice and access principles area concerning because of the self-governance challenges identified (i.e. lack of omnicompetence, time and structure). Nevertheless, these romantic, impractical principles continue to serve as a foundation for ongoing privacy efforts around the world.

In what follows, four examples that draw from the FTC's FIPPs are discussed: the 2012 FTC Data Privacy Report, the White House's Consumer Privacy Bill of Rights, and two proposals from the US Congress. The self-governance concerns highlighted in each example will be identified and critiqued through references to Lippmann and recent privacy research.

The 2012 FTC's Data Privacy Report

In March of 2012, the FTC released *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers* (FTC, 2012). The report was the result of a review process initiated in 2010 after the release of a preliminary report that proposed a framework for protecting consumer privacy in the 21st century.[6] The eventual framework released in the 2012 report articulates best practices for entities involved in data management, and is intended to direct future law and policy efforts (FTC, 2012: vii).

The report has three areas of focus: 1) privacy by design, 2) simplified choice for businesses and consumers, and 3) greater transparency. The 'greater transparency' section will be the focus of the current analysis. In this section, the FTC describes three strategies:

1) Privacy Notices: "Privacy notices should be clearer, shorter, and more standardized to enable better comprehension and comparison of privacy practices."
2) Access to Data: "Companies should provide reasonable access to the consumer data they maintain; the extent of access should be proportionate to the sensitivity of the data and the nature of its use."
3) Consumer Education: "All stakeholders should expand their efforts to educate consumers about commercial data privacy practices." (FTC, 2012: viii)
4) The privacy notice strategy clearly draws from the notice principle, while the data access and consumer education strategies extend from choice and access.

Privacy notices

Though the report acknowledges that privacy policies tend to be too long, complicated, and generally ineffective at informing consumers about data practices, the FTC maintains a strong belief in their value. According to the FTC, privacy notices simply need to be shorter, more concise, and easier to understand. At the same time, the FTC notes that calls for standardized policies are only appropriate for certain notice components, stating, "privacy statements should account for variations in business models across different industry sectors, and prescribing a rigid format for use across all sectors is not appropriate" (FTC, 2012: 62).

Therefore, the FTC's plan requires that users read and understand a variety of privacy notices and policies equal to the variety of entities collecting their data.

What might Walter Lippmann say about this?

Referring to the fallacy of democracy, Lippmann (1927) writes:

> The individual man does not have opinions on all public affairs. He does not know how to direct public affairs. He does not know what is happening, why it is happening, what ought to happen. I cannot imagine how he could know, and there is not the least reason for thinking, as mystical democrats have thought, that the compounding of individual ignorances in masses of people can produce a continuous directing force in public affairs. (p. 29)

In the context of data privacy self-management, Lippmann's argument (i.e. lack of omnicompetence, time and structure) suggests that it is unrealistic to expect that every user, of every age and skill-set, would be able to engage with the evolving TOS and privacy policies of a multitude of data-driven entities. The challenges users face can be broken down into the following areas: 1) TOS and privacy policies are too long and 2) TOS and privacy policies are difficult to understand.

TOS and privacy policies are too long

A seminal study by McDonald and Cranor (2008) assessed privacy policies from the 75 most popular websites on the Internet in 2005. At the time, the policy lengths McDonald and Cranor found ranged from 144 words to 7669 words, with the median being 2500 words. When they asked participants to *skim* a selection of policies, they found that median time to skim one policy ranged from 18 to 26 minutes. They estimated that, at the time, Americans would likely have had to spend 201 hours per year if they read all of the privacy policies they came into contact with (an average of 40 minutes a day). Though the average length in 2005 was 2500 words, as of March 2012, according to the British website *Which?*, the lengths of the terms and conditions for some of the Internet's most popular sites were as follows: Paypal 36,275 words, iTunes 19,972 (2456 privacy, 17,516 terms of use), Facebook 11,195 (6910 privacy, 4285 terms of use), Twitter 4445 and Google 4099 (Parris, 2012). This suggests that TOS and privacy policies have grown in length since 2005, though the current average remains unclear. It should be added that calls for data privacy transparency have contributed to the release of more detailed TOS and privacy policies (to include retention periods, disclosure and sharing policies), as well as transparency reports and law enforcement handbooks (see: Cardozo et al., 2014; Clement and Obar, 2014, 2015a). This suggests that calls for greater transparency will only increase the amount of reading required, making informed consent more of an impossibility. Indeed, it would be nice if every user had the ability, time and interface to achieve informed consent in every context, but this seeming impossibility suggests that the efforts drawn from the notice principle aught to be reconsidered in favor of more pragmatic models that produce results.

TOS and privacy policies are difficult to understand

If users found the time to read TOS and privacy policies, they would still encounter what Solove (2012) refers to as the "privacy self-management consent dilemma" (p. 1883) because policies are difficult to understand (Reidenberg et al., 2014). As professor Paul Ohm, former legal advisor to the FTC once wrote: "Nobody reads privacy policies, and even if people did, they would not be likely to understand them, because they are often very long and full of legalese" (2012: 930). Speaking at a town hall meeting in 2007, then FTC Commissioner Jon Leibowitz described a similar concern:

> Initially, privacy policies seemed like a good idea. But in practice, they often leave
> a lot to be desired. In many cases, consumers don't notice, read, or understand the
> privacy policies. They are often posted inconspicuously via a link at the very bottom of the site's homepage—and filled with fine-print legalese and technotalk. (Leibowitz, 2007: 4)

A recent study by Reidenberg et al. (2014) demonstrates how difficult it is to achieve common understanding—and thus, a consistent informed consent—of privacy policies. The study asks users of varying levels of privacy expertise to evaluate data sharing, retention and deletion policies, and assesses the extent of agreement within and across groups. The findings suggest that within groups of individuals with similar levels of privacy expertise there is considerable disagreement about what policies say. Expert users are identified as having an easier time understanding policies than users in other groups. The authors emphasize that perhaps these findings add to the literature suggesting the failure of the notice and choice framework in general (Reidenberg et al., 2014). These findings contribute to the validation of what is referred to anecdotally as 'the biggest lie on the Internet'—namely, 'I have read and agree to the terms' (Finley, 2012; Greiner, 2012). Access to privacy

policies may be an important first-step, but they still leave users far from achieving data privacy self-management.

Access to data

The FTC organizes their discussion of data access into three sections: a) data for marketing, b) data for eligibility decisions, and c) data for other purposes. In their first section on data for marketing, the FTC notes that calls for complete access and correction rights are unrealistic, mainly because of the prohibitive financial costs for organizations intending to use the data. Instead, the FTC suggests that these organizations provide "a list of the categories of consumer data they hold, and the ability to suppress the use of such data for marketing" (FTC, 2012: 65). Moving away from this straightforward approach they also recommend that entities provide more individualized access or 'granular choices' for opt-in or opt-out when necessary. Here the FTC again begins with an attempt at standardization and simplicity, and then completes their recommendation with guidelines that would ensure variation across entities, adding complexity and work for users.

The second data access category refers to entities that collect data for use by "creditors, employers, insurance companies, landlords, and other entities involved in eligibility decisions" (FTC, 2012: 66). In these instances, the FTC notes that the *Fair Credit Reporting Act* (FCRA) applies. Specifically the provisions that provide consumers with the ability to access and correct all information contained in consumer reports. The FTC recognizes the demands resulting from these guidelines grow increasingly complex, "as more and more consumer data becomes available . . . (including data collected from social media sites) companies are increasingly finding new opportunities to compile, package, and sell that information" (FTC, 2012). For example, the FTC had at the time issued warning letters to an organization that had collected public records and then developed apps to be used by other entities in order to learn information about friends, co-workers, neighbors, or potential suitors. The report added that the applicability of FCRA was unclear in this situation as the specific use of the apps had not been determined. In this instance, the FTC's proposal, again, presents users with an important data access opportunity without explaining how to pragmatically and effectively maintain data privacy self-management amidst an evolving and immense Big Data industry.

The third section of the report refers to entities using data for purposes other than marketing or eligibility. The FTC notes:

> These businesses may encompass a diverse range of industry sectors. They may include businesses selling fraud prevention or risk management services, in order to verify the identities of customers. They may also include general search engines, media publications, or social networking sites. They may include debt collectors trying to collect a debt. They may also include companies collecting data about how likely a consumer is to take his or her medication, for use by health care providers in developing treatment plans. (FTC, 2012: 67)

In each of these instances, the FTC recommends a sliding scale approach related to the sensitivity of the data. When data is more sensitive, and the possibility of damages to the individual is greater, individualized notice, access, and correction rights should be granted. The FTC adds that as a minimum requirement, companies should provide consumers with basic information about the type of data being collected as well as the data sources.

Expanding upon these recommendations, the FTC also singles out data brokers as specific organizations that must do more to increase public awareness of their industry in general and of their data collection and management practices in particular. The FTC recommends that to increase transparency, data brokers must reach out to those they have collected data from (a population of individuals that they generally do not communicate with) and provide those individuals with access to their stockpiles of data. One strategy the FTC mentions is the creation and maintenance of a website where data brokers identify themselves and explain how they collect, organize, re-package, and sell data. The website should also identify the company types that buy data from brokers. The FTC also recommends that brokers use the website to explain access rights and other choices offered to consumers, "and could offer links to their own sites where consumers could exercise such options" (FTC, 2012: 69).

What might Walter Lippmann say about this?

> There is [. . .] nothing particularly new in the disenchantment which the private citizen expresses by not voting at all, by voting only for the head of the ticket, by staying away from the primaries, by not reading speeches and documents, by the whole list of sins of omission for which he is denounced. I shall not denounce him further. My sympathies are with him, for I believe that he has been saddled with an impossible task and that he is asked to practice an unattainable ideal. I find it so myself, for, although public business is my main interest and I give most of my time to watching it, I cannot find time to do what is expected of me in the theory of democracy; that is, to know what is going on and to have an opinion worth

> expressing on every question which confronts a self-governing community. (Lippmann, 1927: 10–11)

In each of the data access scenarios, the FTC offers suggestions that, if followed, will increase the transparency of Big Data. Though it does not create an exhaustive list of all organizations involved, or every potential Big Data practice, the FTC does begin to identify the breadth and depth of the industry. That being said, if Lippmann's omnicompetence, time and structure concerns are applied in this context, it is arguable that the FTC's recommendations for across-the-board increase in communication with consumers, access to data stockpiles, and opportunities for data control, create an impossible scenario for achieving data privacy self-management.

Solove (2012) describes a variety of structural problems that access models like the one proposed by the FTC appear to ignore. Similar to concerns articulated by Lippmann, Solove identifies a 'problem of scale'. There are too many entities, too many quickly moving parts, too many stockpiles and too many data points to expect a consistent, exhaustive and ubiquitous data privacy self-management. The suggestion that users could manage the data collected by one entity like Facebook seems challenging enough—one needs only examine the case of Max Schrems in Germany and the 1200 pages of personal data that he received (O'Brien, 2012)—once the number of data managers is multiplied, the impossibility is amplified. Ben-Shahar and Schneider (2011) describe an "overload effect" specific to the management of complex legal obligations that makes it difficult for individuals to remember, let alone manage complex legal situations. They write:

> Lawmakers have no good solution to this problem. There is rarely a good solution in principle: incomplete disclosure leaves people ignorant, but complete disclosure creates crushing overload problems. (Ben-Shahar and Schneider, 2011: 688)

The FTC's plan, epitomizing romantic and impractical calls for self-governance, proposes that thousands of organizations, large and small, operating in different markets, for different purposes, with different levels of expertise and specialization, using different interfaces, each communicate with consumers, offering them access to an equally diverse set of data stockpiles, privacy, and TOS policies. Furthermore, the FTC's plan currently operates within a national context; Big Data's increasingly global context would greatly expand the complexity of the tasks at hand. The FTC's plan creates potentially millions of pages for consumers to read through, an infinite number of data points to check, understand, critique, and manage.

Consumer education

If that wasn't enough, in its report, the FTC also recommends that organizations involved in data collection, trade or usage, engage in public education about the Big Data industry. Organizations are encouraged to develop and share articles, blog posts, videos, games, etc. that can inform the public about the wide variety of Big Data practices and technologies, as well as about user rights relative to current data management.

What might Walter Lippmann say about this?

> The usual appeal to education can bring only disappointment. For the problems of the modern world appear and change faster than any set of teachers can grasp them, much faster than they can convey their substance to a population of children. If the schools attempt to teach children how to solve the problems of the day, they are bound always to be in arrears. (Lippmann, 1927: 17)

The piling on of impossible tasks in the FTCs proposal is an ideal context for the application of *The Phantom Public*. Walter Lippmann refers to the "fallacy of democracy" as "mystical" (Lippmann, 1927: 28) to suggest that romantic appeals to self-governance convey fantastic possibilities that, while pleasant to the ear, are impossible to achieve in reality. Despite the long history of self-governance critique, and a growing literature calling out flawed notice and choice privacy policy, the FTC and others to be discussed, continue to perpetuate the fallacy of data privacy self-management.

White House Consumer Privacy Bill of Rights

In February of 2012 the White House released *Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy* (White House, 2012). In a letter from President Obama at the beginning of the report, he notes that this proposal is to serve as a blueprint for self-regulation within the Big Data industry, as well for Congressional efforts to fill gaps in existing Federal law.

At the center of the report is a proposed Consumer Privacy Bill of Rights. Of the seven rights,[7] three (first, second and fifth) relate directly to data privacy self-management, drawing from the Fair Information Privacy Principles the FTC developed in the 1970s. The first right presented is: "Individual Control, consumers have a right to exercise control over what personal data companies collect from them and how they

use it" (White House, 2012: 11). The report goes on to say that at the time of collection, entities collecting data must provide users with choices about data collection, use, disclosure, and sharing that are relevant to the scale, scope and sensitivity of the personal data in question.

> For example, companies that have access to significant portions of individuals' Internet usage histories, such as search engines, ad networks, and online social networks, can build detailed profiles of individual behavior over time. These profiles may be broad in scope and large in scale, and they may contain sensitive information, such as personal health or financial data. In these cases, choice mechanisms that are simple and prominent and offer fine-grained control of personal data use and disclosure may be appropriate. (White House, 2012)

The right to individual control also guarantees that users should have access to innovative technologies that can help ensure user control over data. Companies involved in online data collection should ensure that their products have detailed privacy settings that users can modify as appropriate to control what data is collected and when. Going a step further, the report notes that applications that enable 'Do Not Track' functionalities should be built-into these products to further strengthen user control by offering the opportunity to opt-out of data collection and use, especially by third parties.

The right of individual control also extends to data collected and used by data brokers. The White House acknowledges that this poses considerable challenges because data brokers are removed from direct interactions with the general public, and because many individuals are unaware of the data aggregation industry. Similar to the FTC report, the White House proposal encourages data brokers to develop mechanisms for increasing consumer access and control over personal data, and also recommends that data brokers engage in public awareness initiatives.

The second right is: "Transparency, consumers have a right to easily understandable and accessible information about privacy and security practices" (White House, 2012: 14). The Administration notes that plain language statements about data collection, use disclosure and retention practices should be well-integrated into all technologies that produce data. Not only should there be a blanket statement associated with each technology, but organizations should communicate with citizens about data privacy concerns "when they are most relevant." This suggests that users should be given access to data privacy information or consent materials at multiple points during an interaction as new actions lead to new data. These responsibilities also apply to data brokers.

The fifth right is: "Access and Accuracy, consumers have a right to access and correct personal data in usable formats, in a manner that is appropriate to the sensitivity of the data and the risk of adverse consequences to consumers if the data is inaccurate" (White House, 2012: 19). Similar to concerns articulated by the FTC, this right emphasizes that access to data being used for eligibility decisions should be guaranteed. Specifically, sectors whose data management practices are not already covered by Federal privacy laws such as the *Health Insurance Portability and Accountability Act* and FCRA are the focus. To ensure protections, entities should not only provide user access to data, but the ability to correct, delete, or suppress data where appropriate. The White House recently released a draft of the *Consumer Privacy Bill of Rights Act of 2015* (White House, 2015) with much of the same language discussed included in the proposal.

Data privacy self-management and recent congressional efforts

In the United States, a variety of recent legislative efforts have included similar provisions that also perpetuate the fallacy of data privacy self-management. One example was the *Do Not Track Kids Act of 2013* (Markey et al., 2013—not enacted). Among the components of the bill was an 'openness' principle which included a variety of notice, choice and access provisions, but geared towards minors. The bill stated that those operating websites targeted to minors or involved in data collection of minors must:

> provide each minor using the website, online service, online application, or mobile application [...] with a clear and prominent means [...] to obtain any personal information of the minor that is in the possession of the operator from the operator [...] to challenge the accuracy of personal information of the minor that is in the possession of the operator; and [...] if the minor establishes the inaccuracy of personal information [...] to have such information erased, corrected, completed, or otherwise amended. (pp. 18–20)

This bill epitomizes the inability of regulators to see beyond the fallacy. To expect that minors are capable of addressing all of the challenges identified herein, suggests that members of the US Congress continue to propose bills with romantic notions of citizen empowerment in mind, without focusing on pragmatics. The *Data Broker Accountability and Transparency Act of 2015* (Markey et al., 2015), currently being reviewed by Congress, includes a number of notice and choice

provisions similar to those presented by both the FTC and the White House. For example, there is the requirement that, "a covered data broker shall maintain an Internet website and place a clear and conspicuous notice on that Internet website instructing an individual how [...] to review information under subsection (b)(1) (personal information)". This requirement demonstrates once again that regulatory efforts continue to draw from the problematic FIPPs.

Both of these legislative efforts champion the notion of data privacy self-management as outlined in greater detail in both the FTC and White House reports. These efforts call for individuals to have access to their data, and the opportunity to understand and control how their data is being collected, managed and used. Lacking from each of these proposals however are pragmatics designed to ensure the realization of data privacy self-management.

Discussion

> I have been reading the new recommendations used to encourage data privacy self-management in the United States. After reading them I do not see how any one can escape the conclusion that the digital citizen must have the appetite of a data miner and infinite time at their disposal. In these new proposals the digital citizen studies the problems of Big Data, and not the structural detail. They are told in reports from the FTC, the White House and the U.S. Congress about reading terms of service agreements, privacy policies, the seemingly impossible task of communicating with potentially thousands of companies of all shapes, sizes and abilities, including data brokers who have never been in the business of communicating with the public, reviewing data sets even more diverse and complex, and being asked to critically engage with this data and its management, as well as to understand the financial strategies of the entities wishing to profit from this data. After completing all of this, the digital citizen is then asked to provide restrictions for each of these data-driven entities, continuously imposing multiple unique limitations on a constantly and quickly evolving limitless number of stockpiles and applications. This is all accomplished through the myriad of interfaces determined appropriate by the data-driven entities that will be (and you can be sure of this) more than happy to welcome the millions of assuredly well-behaved and well-informed information auditors into their vaults.
>
> But nowhere in these well-meant recommendations is the sovereign digital citizen of the future given a hint as to how, while earning a living, rearing children and enjoying life, they are to keep informed about the progress of this swarming confusion of problems. Furthermore, the authors of these guidelines have missed a decisive fact: the digital citizen gives but a little of their time to these affairs, has but a casual interest in Big Data and but a poor appetite for theory. It never occurs to these preceptors of data privacy self-management to provide the student with the rule by which they can know whether on Thursday it is their duty to consider Google's new data points or Facebook's new privacy policy, nor how, if they determine on Thursday to express their sovereign will on the data point question, the digital citizen is to repair those gaps in their knowledge of that question which are due to having been preoccupied the day before in expressing their sovereign will about Walmart's new data management system or some data broker's mission statement.
>
> Yet the digital citizen cannot know all about everything all the time, and while they are watching one thing a thousand others undergo great changes. Unless the digital citizen can discover some rational ground for fixing attention where it will do the most good, and in a way that suits inherently amateurish equipment, the digital citizen will be as bewildered as a puppy trying to lick three bones at once.

Once again, had Walter Lippmann been concerned with the nature of our digital existence, perhaps this is how the second chapter of *The Phantom Public* entitled "The Unattainable Ideal," might have started.

What is missing from the detailed guidelines formulated by the FTC, the White House and the US Congress are pragmatic strategies for data control and protection that take into consideration all of the limitations and challenges articulated by Lippmann and others. We have only begun to realize the challenges associated with the evolving universe of Big Data and new, tangential challenges are quickly coming into view. For example, at the University of Toronto, IXmaps researchers are beginning to uncover political economies of data packet transmission. For instance accessing a website or sending an email could result in routing variegations, sometimes with data crossing national boundaries, presenting data collection and surveillance concerns (Obar and Clement, 2012). Imagine having to understand, manage and control, not only the myriad data stockpiles that exist, but also the routing data associated with every data transmission.

> A partial answer to the fallacy of data privacy self-management is articulated by Lippmann:
>
> The interest of the public is not in the rules and contracts and customs themselves but in the maintenance of a régime of rule, contract and custom. The public is interested in law, not in the laws; in the method of law, not in the substance; in the sanctity of contract, not in a

> particular contract; in understanding based on custom, not in this custom or that. (Lippmann, 1927: 95)

Even if we had the faculties *and* the system for data privacy self-management, the digital citizen has little time for data governance. What we desire is the freedom to pursue the ends of digital production, without being inhibited by the means. The average digital citizen wants privacy, and safety, but cannot complete all that is required for its protection.

If it is true that the fallacy of democracy is similar to the fallacy of data privacy self-management, then perhaps the imperfect yet pragmatic answer to the fallacy of data privacy self-management is to similarly introduce a system of representative governance. *Representative data management* could contribute to the protection of personal data while freeing individuals from the impossible task of data privacy self-management. As scholars discuss models of privacy-focused infomediaries (e.g. Mowbray et al., 2012), banks and organizations like the US-based Lifelock are already offering identity theft protection services to address a burgeoning consumer demand. Other more robust commercial and non-commercial representative data managers should be developed (perhaps modeled after accounting services), whose task will be to manage our massive array of digital dossiers. The accountant analogy fits well as it provides an example of an infomediary that interprets a complicated set of systems (i.e. the tax code and government tax systems) as well as our own personal financial data, while interfacing with the general public in a user-friendly manner. Accountants also engage in a form of reputation management on our behalf, helping consumers appropriately represent their financial status to the government. Each of these elements of the consumer–accountant relationship should be modeled as representative data managers develop to meet consumer demand. Though generating demand seems a difficult task (would you pay someone to manage your online photos?) consumer demand for representative data managers aught to increase, in light of the myriad Big Data products and services increasingly driving eligibility decision-making (e.g. Pasquale, 2015). These data managers would be responsible for continuously collecting and patrolling our Big Data, while offering individuals a simplified, all-encompassing interface that can be managed and controlled from afar, perhaps annually. Non-commercial options should be developed, to offer services that meet diverse consumer demands, in light of growing concerns over digital forms of discrimination (Gangadharan et al., 2015; Pasquale, 2015). Future research should address the extent to which representative data management can address the challenges identified by the fallacy of data privacy self-management, as well as methods for generating demand for services.

One issue to be clarified when considering the application of Lippmann's self-governance critique and the suggestion of a representative solution is the imperfect analogy linking the fallacy of democracy and the fallacy of data privacy self-management. Indeed, one does not mirror the other perfectly. For example, democratic government, whether popular or representative, aims to engender a system of governance for both the individual and the society as a whole. It would be a stretch to suggest that calls for data privacy self-management make similar claims to governance of the entire Big Data universe. What matters in terms of the application of the analogy is the analysis of impractical self-governance proposals and the long history of self-governance critique that led to the Dewey–Lippmann debate, and should now extend to the Big Data context.

As we dip our heads beneath the waterline and force the Big Data iceberg into public view, pragmatic, realistic approaches to transparency, privacy and control are required. If we cling to romantic fallacy, we perpetuate what Lippmann referred to as "a false ideal." As Lippmann states: "I do not mean an undesireable ideal. I mean an unattainable ideal" (1927: 29). In the Big Data context, the challenge of citizen empowerment described here should be seen as an extension of the longstanding self-governance debate, exemplified by the works of John Dewey and Walter Lippmann. Both framed their arguments in opposition to what they viewed as a system struggling to find an autonomous and efficacious role for the citizen. In the Big Data context, Lippmann's pragmatism champions citizen empowerment by critiquing a governance model that fails to achieve practical self-governance outcomes. Achieving pragmatic alternatives to data privacy self-governance will not be easy; however, the first step towards a plan that expresses the true possibilities of its subject requires movement beyond romantic notions that remain as impossible as direct democracy within a nation of millions.

Declaration of conflicting interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. In this instance, the term "digital citizen" refers to any individual who engages with digital media technologies, no matter how extensively. This broad definition is necessary to emphasize that light users also generate data.

2. The "biggest lie on the internet" refers to the statement "I have read and agree to the terms" (Finley, 2012; Greiner, 2012).
3. Lippmann doesn't use the phrase "fatal flaws," which is used herein to organize the most relevant self-governance concerns identified in *The Phantom Public*.
4. Though the classical direct democracy of ancient Greece is often held up as a normative ideal, in actuality, the Greek political system was prejudiced, sexist and emphasized exclusion in many respects.
5. The Canadian privacy law most relevant is the Personal Information and Electronic Documents Act (PIPEDA). In Europe, the privacy law most relevant is the EU Data Protection Directive.
6. See: www.ftc.gov/os/2010/12/101201privacyreport.pdf
7. The White House's Consumer Privacy Bill of Rights includes the following seven rights: 1) individual control, 2) transparency, 3) respect for context, 4) security, 5) access and accuracy, 6) focused collection, and 7) accountability.

References

Acquisti A, Gross R and Stutzman F (2014) Face recognition and privacy in the age of augmented reality. *Journal of Privacy and Confidentiality* 6(2): 1–20.

Bamford J (2012) The black box: Inside America's massive new surveillance center. *Wired.co.uk*, 30 March. Available at: www.wired.co.uk/magazine/archive/2012/05/features/the-black-box (accessed 18 December 2014).

Ben-Shahar O and Schneider CE (2011) The failure of mandated disclosure. *University of Pennsylvania Law Review* ▮▮: 647–749.

Burn-Murdoch J (20 May 2013) Critics of big data have overlooked the speed factor. In: The Guardian. Available at: http://www.theguardian.com/news/datablog/2013/may/20/big-data-critics-overlooked-speed-not-size (accessed 5 August 2015).

Cardozo N, Cohn C, Higgins P, et al. (2014) Who has your back? Protecting your data from government requests. Available at: https://www.eff.org/files/2014/05/15/who-has-your-back-2014-govt-data-(accessed 21 May 2015).

Carey JW (1987) The press and the public discourse. *Center Magazine* 20(2): 4–16.

Citron DK (2009) Cyber civil rights. *Boston University Law Review* 89: 61–125.

Citron DK and Franks MA (2014) Criminalizing revenge porn. *Wake Forest Law Review* 49: 345.

Clement A and Obar JA (2014) *Keeping internet users in the know or in the dark: An analysis of the data privacy transparency of Canadian internet service providers*. IXmaps report, 27 March. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2491847 (accessed 21 May 2015).

Clement A and Obar JA (2015a) *Keeping internet users in the know or in the dark: An analysis of the data privacy transparency of Canadian internet carriers*. IXmaps 2014 report. Available at: http://ixmaps.ca/transparency-2014.php (accessed 21 May 2015).

Clement A and Obar JA (2015b) Canadian internet "boomerang" traffic and mass NSA surveillance: Responding to privacy and network sovereignty challenges. In: Geist M (ed.) *Law, Privacy and Surveillance in Canada in the Post-Snowden Era*. Ottawa, ON: University of Ottawa Press.

Constine J (2012) How big is Facebook's data? 2.5 billion pieces of content and 500+ terabytes ingested every day. *TechCrunch*, 22 August. Available at: http://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces- of-content-and-500-terabytes-ingested-every-day (accessed 18 December 2014).

Couts A (2012) Meet the online snoops selling your dirty laundry and how you can stop them. In: Digital Trends. Available at: http://www.digitaltrends.com/social-media/unmasking-the-data-brokers-who-they-are-and-what-they-know-about-you (accessed 18 December 2014).

Dewey J (1925) Practical democracy. *The New Republic* 45: 52–54.

Dewey J (1927; 2012) *The Public and Its Problems: An Essay in Political Inquiry*. University Park, PA: Penn State Press.

Federal Trade Commission (2012) *Protecting consumer privacy in an era of rapid change: Recommendations for businesses and policymakers*. FTC Report, March 2012. Available at: www.ftc.gov/os/2012/03/120326privacyreport.pdf (accessed 18 December 2014).

Federal Trade Commission (2014) Data brokers: A call for transparency and accountability. Report, May 2014. Available at: https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf (accessed 21 May 2015).

Finley K (2012) Putting an end to the biggest lie on the Internet. *Techcrunch*. Available at: http://techcrunch.com/2012/08/13/putting-an-end-to-the-biggest-lie-on-the-internet (accessed 18 December 2014).

Gangadharan S, Eubanks V and Barocas S (eds) (2015) Data and discrimination: Collected essays. *New America Foundation*. Available at: https://www.newamerica.org/downloads/OTI-Data-an-Discrimination-FINAL-small.pdf (accessed 5 August 2015).

Greiner L (2012) 'I have read and agree to the terms': The web's biggest lie? In: The global and mail, technology. Available at: http://www.theglobeandmail.com/report-on-business/small-business/sb-digital/biz-categories-technology/i-have-read-and-agree-to-the-terms-the-webs-biggest-lie/article4642898 (accessed 18 December 2014).

Hargittai E and Hinnant A (2008) Digital inequality differences in young adults' use of the Internet. *Communication Research* 35(5): 602–621.

Howes TA (15 June 2015) Spark speeds towards the next data processing revolution. In: IBM blogs. Available at: http://www.ibmbigdatahub.com/blog/spark-speeds-towards-next-data-processing-revolution (accessed 5 August 2015).

IBM (2015a) The four v's of Big Data. *Infographic*. Available at: http://www.ibmbigdatahub.com/infographic/four-vs-big-data (accessed 21 May 2015).

IBM (2015b) Why speed matters for big data and analytics. *Infographic*. Available at: http: //www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=ST&infotype=SA&appname=STGE_NI_EZ_USEN&htmlfid=NIJ1

2345USEN&attachment=NIJ12345USEN.PDF (accessed 5 August 2015).

Leibowitz J (2007) So private, so public: Individuals, the internet and the paradox of behavioral Marketing. *Federal Trade Commission*. Available at: https://www.ftc.gov/sites/default/files/documents/public_statements/so-private-so-public-individuals-internet-paradox-behavioral-marketing/071031ehavior_0.pdf (accessed 21 May 2015).

Lippmann W (1927; 2009 edition) *The Phantom Public*. New Jersey: Transaction Publishers.

McDonald AM and Cranor LF (2008) The cost of reading privacy policies. *I/S: A Journal of Law and Policy* 4(3): 540–565.

Manyika J, Chui M, Brown B, et al. (2011) *Big data: The next frontier for innovation, competition, and productivity*. Report for McKinsey Global Institute, May 2011. Available at: http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation (accessed 18 December 2014).

Markey, Kirk and Blumenthal (2013) Do not track kids act of 2013. In: 113th Congress, 1st Session, 13 November 2013. Available at: http://www.gpo.gov/fdsys/pkg/BILLS-113s1700is/pdf/BILLS-113s1700is.pdf (accessed 21 May 2015).

Markey, Blumenthal, Whitehouse, et al. (2015) Data Broker Accountability and Transparency Act of 2015. In: 114th Congress, 1st Session. 4 March 2015. Available at: http://www.gpo.gov/fdsys/pkg/BILLS-114s668is/pdf/BILLS-114s668is.pdf (accessed 21 May 2015).

Microsoft (2015) Improving the Microsoft services agreement and privacy statement for consumers. In: The Official Microsoft Blog. Available at: http://blogs.microsoft.com/blog/2015/06/04/improving-the-microsoft-services-agreement-and-privacy-statement-for-consumers/ (accessed 5 August 2015).

Mowbray M, Pearson S and Shen Y (2012) Enhancing privacy in cloud computing via policy-based obfuscation. *The Journal of Supercomputing* 61(2): 267–291.

Napoli PM and Obar JA (2014a) The emerging mobile Internet underclass: A critique of mobile Internet access. *The Information Society* 30(5): 323–334.

Napoli PM and Obar JA (2014b) The mobile conversion, Internet regression, and the "re-passification" of the media audience (21 October 2014). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2511393 (accessed 5 August 2015).

Noamgalai.com (2015) Noam Galai photography. Available at: http://noamgalai.com/scream.htm (accessed 5 August 2015).

Obar JA and Clement A (2012) Internet surveillance and boomerang routing: A call for Canadian network sovereignty. In: *TEM 2013: Proceedings of the Technology and Emerging Media Track-Annual Conference of the Canadian Communication Association*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2311792 (accessed 21 May 2015).

O'Brien KJ (5 February 2012) Austrian law student faces down Facebook. *The New York Times*. Available at: http://nyti.ms/1DMSs8P (accessed 21 May 2015).

Ohm P (2012) Branding privacy. *Minnesota Law Review* 97: 907–989.

Parris R (2012) Online TandCs longer than Shakespeare plays—Who reads them? *Which, Conversation*. Available at: http://conversation.which.co.uk/technology/length-of-website-terms-and-conditions (accessed 18 December 2014).

Parsons C (6 March 2014) The murky state of Canadian telecommunications surveillance. Available at: https://citizenlab.org/2014/03/murky-state-canadian-telecommunications-surveillance/ (accessed 21 May 2015).

Pasquale F (2015) *The Black Box Society*. Cambridge, MA: Harvard University Press.

Prieger JE (2013) The broadband digital divide and the economic benefits of mobile broadband for rural areas. *Telecommunications Policy* 37(6): 483–502.

Reese C (23 July 2014) How speed is going to change the data game. In: IBM blogs. Available at: http://www.ibmbigdatahub.com/blog/how-speed-going-change-data-game (accessed 5 Aug 2015).

Reidenberg JR, Breaux T, Cranor LF, et al. (2014). Disagreeable privacy policies: Mismatches between meaning and users' understanding. Fordham law legal studies research paper. Available at: http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2418297 (accessed 21 May 2015).

Schneier B (2015) *Data and Goliath: The Hidden Battles to Capture Your Data and Control Your World*. New York, NY: WW Norton and Company.

Schudson M (2008) The "Lippmann-Dewey debate" and the invention of Walter Lippmann as an anti-democrat 1985–1996. *International Journal of Communication* 2: 1031–1042.

Solove DJ (2002) Identity theft, privacy, and the architecture of vulnerability. *Hastings Law Journal* 54: 1227.

Solove DJ (2012) Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review* 126: 1880–1903.

Swire P (2012) Why the Federal government should have a privacy policy office. *Journal on Telecommunication and High Technology Law* 10: 41–51.

The economist (25 February 2010) Data, data everywhere. *The Economist*. Available at: http://www.economist.com/node/15557443 (accessed 21 May 2015).

Wayne LD (2012) The data-broker threat: Proposing Federal legislation to protect post-expungement privacy. *The Journal of Criminal Law and Criminology* 102(1): 253–282.

Whipple M (2005) The Dewey-Lippmann debate today: Communication distortions, reflective agency, and participatory democracy. *Sociological Theory* 23(2): 156–178.

White House (2012) Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global digital economy. February 2012. Available at: http://www.whitehouse.gov/sites/default/files/privacy-final.pdf (accessed 18 December 2014).

White House (2015) Administration discussion draft: Consumer Privacy Bill of Rights Act of 2015. Available at: https://www.whitehouse.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf (accessed 21 May 2015).

Wray R (2009) Internet net heads for 500bn gigabytes. *The Guardian*, 18 May. Available at: http://gu.com/p/27na9 (accessed 18 December 2014).

5