Original Research Article

Big Data & Society
July–December 2014: 1–11
© The Author(s) 2014
DOI: 10.1177/2053951714564228
bds.sagepub.com

How web tracking changes user agency in the age of Big Data: The used user

Sylvia E Peacock

Abstract

Big Data enhances the possibilities for storing personal data extracted from social media and web search on an unprecedented scale. This paper draws on the political economy of information which explains why the online industry fails to self-regulate, resulting in increasingly insidious web-tracking technologies. Content analysis of historical blogs and request for comments on HTTP cookies published by the Internet Engineering Task Force illustrates how cookie technology was introduced in the mid-1990s, amid stark warnings about increased system vulnerabilities and deceptive personal data extractions. In conclusion, online users today are left with few alternatives but to enter into unconscionable contracts about the extraction of their personal data when using the Internet for private purposes.

Keywords

Web-tracking technology, HTTP cookies, personal data, agency, political economy of information, Big Data

Introduction

Web tracking happens across platforms; it is the unseen and unauthorised extraction, storage, analysing, selling, buying and auctioning of personal online data appropriated by one or more remote online corporate actors. Apart from appearing in recent patent applications (e.g. Parreira, 2013), synonyms include ‘Internet tracking’ (Cumbley and Church, 2014) or ‘online tracking’ (Knight and Saxby, 2014). In the industry, the definitions offered inform the online user that she is followed through a network of sites while her online activities are recorded (e.g. Laredo Group Inc. n.d.; Office of the Privacy Commissioner of Canada, 2012).

Big social data and web tracking are intricately connected and their development appears mutually dependent. Without data storage capacities reaching far beyond the exabyte point (Kambatla et al., 2014), web tracking might be no more than an interesting application. But as the *Internet of Things* is looming, it seems as if demand for data storage is continuously increasing (Crump and Harwood, 2014). With this in mind, it seems odd that so little attention is paid to the online user and how she is affected by web tracking. This study focuses on the ‘paradigmatic shift’ towards computational social science (Chang et al., 2014) right down to the level of the individual, namely the status of the online user. Given the increased corporate interest to put IP addresses on myriad other household devices (Crump and Harwood, 2014), it is an appropriate starting point.

All online entries of people who use social media or web search are extracted and stored in Big Data warehouses (Chen et al., 2014; Doctorow, 2008; Lu et al., 2014). Some of the data are used for behaviourally targeted advertising, some for real-time-bidding (Boston Consulting Group, 2012; Weide, 2011). These theoretically limitless personal data ‘teach’ algorithms to increase economic transactions with online agents (Lu et al., 2014; Shroff, 2013). Online agents have no possibilities to limit the extraction, storage and use of their personal data (Acar et al., 2014; Chen et al., 2014); on the contrary, digital footprints containing personal data are constantly growing, due to ongoing advances

Department of Social Science, York University, Toronto, Canada

Corresponding author:
Sylvia E Peacock, York University, 4700 Keele St, Ross Building South, Toronto, ON M3J 1P3, Canada.
Email: speacock@yorku.ca

in big social data extractions, storage and management (Lanier, 2013).

Personal data include all online blogs, pictures, texts, Tweets, emails, videos and technical details attached. Data points contain metadata as well as unstructured and finely granular information, i.e. emotional expressions or affective exchanges (Dwoskin, 2014). Although the term personal data signifies individualised information in an electronic format, their personal nature does not shield them from commodification (Gandy, 1993). For example, people's information about their leisure activities, social networking activities and Internet usage are bought and sold in the consumer data broker industry (Roderick, 2014: 732).

The scope of my analysis covers North America with illustrative empirical material from Canada. For practical purposes, the scope is much wider. Online US business practices set an international precedent on the Internet that seems difficult to reverse. What is more, most governments' desires for people's online information seem well served by the weak or non-existing policies for the treatment of personal data (El Akkad, 2014).[1] The discussion here revolves around web applications like search engines and social media, but they extend beyond this. Many online retailers for goods and services (from booksellers to online smut) increase profits through web tracking and the storage and sales of online customers' personal data (Lanier, 2013; Wondracek et al., 2010).

As a working thesis, I propose that online users do not use the Internet to 'donate' personal data to unknown corporate entities. This may be contentious, as some people might enjoy being targeted by adverts to inform them about consumer goods that fit their personal consumer profile (Turow, 2012). A case can be made, though, for an inherent preference amongst individuals to control the extraction and distribution of personal data. People prefer to be the experts of their own situation (Benello, 1981; Bourdieu, 1990).

As an ideal typical social construct, I propose to view web applications like social media or web search as an online public sphere, a virtual space to advance, discuss, elaborate or search for new personal ideas. Theoretically, social media and web search can bring together a large number of people in a many-to-many discussion and offer myriad platforms to facilitate online exchanges of ideas for mutual benefits. A public sphere, as Habermas defines it, is '[...] a forum in which the private people come together to form a public [and] read[y] themselves to compel public authority to legitimate itself before public opinion' (1989: 25f).

Of course, not just discursive but also deviant behaviour can thrive in places that seem to offer anonymity, for example, the so-called 'trolling' or 'flaming'. But while online deviance is a galling reminder of a breakdown in communication, civility and maturity are not prerequisites for a public sphere. Therefore, deviant online communication does nothing to diminish the notion that social media and web search can serve as an online public sphere. The online public sphere serves as a counter position to the reductive notion of viewing the Internet as an online marketplace.

In theory, a public sphere like the Internet is open to all and non-exclusionary, but the way people make use of it varies in quantity and quality. For example, roughly 83% of the non-institutionalized Canadian population logs on to the Internet at least once a year (Statistics Canada, 2013). So it seems almost everyone is online. Now, although online social networks and web search continue to engage a growing number of users, with 93% of them accessing email, the latter continues to be the predominant reason for people to use the Internet (Korupp, 2006; Korupp et al., 2006; Peacock and Kühnemund, 2007; Statistics Canada, 2011, 2013). Likewise, worldwide diffusion rates differ vastly. In 2014, 40% of the world population uses the Internet, but individual country estimates range between 20% and 80% (International Telecommunication Union (ITU), 2014).[2] So the current focus on the use of web search and social media captures the situation of a subgroup of online users, although it is an increasingly important one (Minister of Industry, 2010: 14; Peacock, 2014; Statistics Canada, 2013). Web search and social media have opened up additional channels of social interaction including new online friendships, companionship and a fresh sense of belonging. Or so it seems.

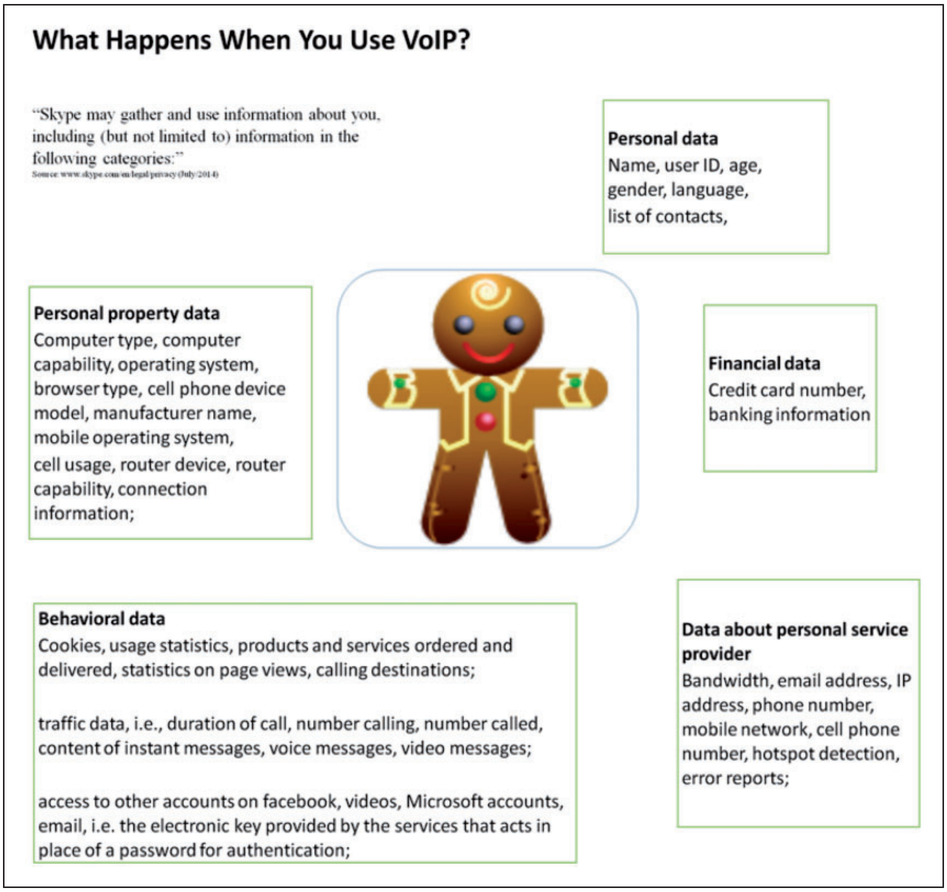
A closer look at social media and web search sharply narrows one's first impression of online diversity. Three of the five most visited online social platforms are owned by two large corporations, Google Inc. and Facebook Inc. The source on which my website ranking is based – alexa.com – is owned by another Internet giant, Amazon.com, that started as an online bookseller. Not accidentally, all these large online corporations are deeply involved in Big Data, more specifically, in big *social* data technology (Lanier, 2013). Already large online businesses are vying for market domination, and with mergers, affiliations and acquisitions ongoing, an ever smaller number of corporations are capturing an increasing amount of people's personal data.

I focus on two questions: firstly, what is the extent of self-regulation that an online user may expect from an online information industry involved in web-tracking technologies? Secondly, how does web tracking affect online user agency? Together with the quasi-monopolisation of social media, non-regulation in online information markets has profound impacts on well-known symptoms of market failure, discussed in the following section. The section thereafter includes a content analysis of the past and current public online subgroup discussions from the Internet Engineering Task Force (IETF) to illustrate the controversial introduction of HTTP cookie technology in the 1990s. I include descriptions of newer web-tracking technologies and how they overcome the defences employed by online users. In the final section, I summarise my most important results and discuss the uncomfortable agreement online agents currently enter into, whenever they use web search or social media.

Theoretical background: Market failure

Companies like Acxiom, Seisint, Datalogix or Epsilon collect, analyse and sell people's personal financial data to profile credit worthiness or calculate people's credit scores (for a critical discussion, see O'Harrow, 2005 or Roderick, 2014). They are highly profitable because accurate information on people's net wealth is at the heart of economic risk reduction. Although personal data are of a more emotional or affective nature and only indirectly offer insights into people's market position, they are now collected in large quantities to inform companies about possible future economic opportunities (Boston Consulting Group, 2012; Numan and DiDomenico, 2013). These future economic possibilities operate in the framework of economic risk reduction, too, because consumer products are determined by data mining results based on personal data. Of course, from a utilitarian view, a certain inconsistency still persists because of the sheer quantity of data extracted, but that seems resolved by current developments in Big Data. To exemplify a number of data points collected, Figure 1 shows what the approximately 31 million users of a well-known voice-over-Internet protocol (VoIP) service agree to when using their service.[3]

Most of the data points stored are affective and individualised – online behaviour, messages, talk content, and the technologies owned and services accessed by the user to connect to VoIP or while her system is



What Happens When You Use VoIP?

"Skype may gather and use information about you, including (but not limited to) information in the following categories:"

Source: www.skype.com/en/legal/privacy (July/2014)

Personal data
Name, user ID, age, gender, language, list of contacts,

Personal property data
Computer type, computer capability, operating system, browser type, cell phone device model, manufacturer name, mobile operating system, cell usage, router device, router capability, connection information;

Financial data
Credit card number, banking information

Behavioral data
Cookies, usage statistics, products and services ordered and delivered, statistics on page views, calling destinations;

traffic data, i.e., duration of call, number calling, number called, content of instant messages, voice messages, video messages;

access to other accounts on facebook, videos, Microsoft accounts, email, i.e. the electronic key provided by the services that acts in place of a password for authentication;

Data about personal service provider
Bandwidth, email address, IP address, phone number, mobile network, cell phone number, hotspot detection, error reports;

Figure 1. Detailed personal data profile accessed by voice-over-IP service ('Skype'). Image credits for gingerbread-man: veryicon.com.

idling. Some of the data cover the details transmitted to third-party Internet service providers and other services the user contacts. In sum, most of the data extracted, stored, analysed or otherwise captured by Skype, – or rather Microsoft[4] – do not include financial information; they depict a unique private person during her private social interactions.[5]

Whereas it once was the state that was interested in finding out what people were up to in their private homes, today it seems that the business of personal data extraction is firmly in the hands of publicly unaccountable corporations (Bamford, 1982; Denardis, 2014; Jussawalla and Cheah, 1987). Access, retention and analyses of all bits and bytes of personal electronic data are becoming increasingly sophisticated and helped along by current developments in Big Data (Acar et al., 2014; Krishnamurthy et al., 2007; Soltani et al., 2009; van Eijk et al., 2012). The built-in trade-off between offering online content and hoarding personal data, the 'convenience versus privacy' exchange, seems firmly resolved in favour of the former (Dumas, 2012: 217f). Industry representatives call this a self-regulated outcome (Dusseault, 2013: 5). It is the online user of social media and web search who bears the burden of their own data extraction, the lack of meaningful alternatives and very few regulations.

Most industrialised countries have a 'marketplace solution' for personal online data protection, with very little regulatory interference due to a dominant belief in a neoliberal economic paradigm (Denardis, 2014: 53ff). As current wisdom has it, the market is the best supplier of online goods and services because it allegedly follows mandates of efficient supply and demand.[6] But notable exceptions exist as to how well a market can distribute goods and services, and one of them is the distribution of information (Baker, 2002). Regulatory inactivity in information markets leads to market failure, partly due to the nature of the information production process (Gandy, 1993). One of the troubles with the current market solution is that information is neither measurable as a unit nor is it priced as an item (Babe, 1983). Of course, there is a price for information – where there is a buyer for people's personal data there will be a price – but essentially it is the production of information that is valued at a price, not immaterial information itself (Babe, 1983).

Other problems are that information is intangible, non-exclusive and has important public good characteristics (Curran, 1991; Gandy, 1993). Production and dissemination cannot be weighed and information is easily carried around as well as effortlessly distributed. Theoretically, information is inexhaustible and if redistributed or sold, still remains with the original seller. These characteristics of nonexcludability and nonrivalry are important aspects of public goods (Baker, 2002; Gandy, 1993). The benefits of public goods to society go by the numbers, and websites on the World Wide Web might illustrate a case in point. Most of them are accessible without a price and as more people produce and share information virtually everybody becomes better off: Information distribution on the Internet is a good example of a Pareto improvement.[7]

Equally problematic and often forgotten is the salience of content, what is *inside* the actual information. Babe challenges the correctness of looking at information only in terms of a unit price, an approach he calls 'at best simplistic and partial' (1983: 132). Instead of looking at information as a commodity, he suggests conceiving it as an *essence* (Babe, 1983). The *essence* of information is defined as the quality or importance of a message. If essence was considered, the extraction of personal data would need to include an ethical framework and considerations about personal sovereignty and integrity, and it could run counter to people's civil rights.

The information industry is very much aware of the often priceless qualitative contents of information. Whenever the invisible extraction of online personal data is detected in connection with an abuse of its essence, opposition forms, corporate statements are given and damage control enacted.[8] Despite this, the information industry continues to treat personal data as goods. As it stands, anonymous Internet users are not lucrative and personal data have a unit price. The mere existence of a price, though, is not the sign of a functioning market. Below, several symptoms are discussed that point towards current market failure.

Large online social media and web search companies are benefitting from their quasi-monopoly status, and time is on their side. Depending on the number of written entries users make on a social media platform, it might constitute quite a feat to migrate their personal data to a future competitor. It is highly questionable, too, whether their migration to a different platform leads to the removal of their personal data in the big database of the previous site. Consequently, the online user is the sole price taker who has few alternatives but to agree to her personal data extraction.[9] The price is set, there is no alternative.[10] The absence of the possibility to walk away from a market exchange indicates a dysfunction in the market.

What is more, there is no equitable allocation of negative effects between buyers and sellers. Some examples of negative effects produced by the current market-led approach are a loss of autonomy, expertise, integrity, developmental possibilities, personal growth, public debate, search for new information and the underuse of online capacities for personal exchange.

Economists tend to use the generic term of *externalities* to label these negative effects (Jussawalla and Cheah, 1987). The impossibility of online users to limit web tracking is symptomatic: 'If externalities cannot be limited, it is a sign of market failure' (Gandy, 1993: 206).

Another indication is the unethical use of regulatory loopholes to gain profitable advantages (Furubotn and Pejovich, 1972). Penalties and rewards matter in order to explain the current incentives for the tracking and storing of personal data. People as well as institutions choose between a set of 'sanctioned behavioural relations' when they extract personal data from online users to gain profitable advantages (Furubotn and Pejovich, 1972: 1131). Currently, incentives for transparent, limited and consensual personal data extractions are low, while profits for invisible web tracking and unlimited data storage are high, all the while costs for storage are decreasing.

Within a decade, corporate access to personal online data has morphed into an economic advantage. Influential economic interest groups are celebrating the increased access to personal data as the creation of a new asset class, indeed, as the 'new oil' (Boston Consulting Group, 2012; World Economic Forum, 2011). These current conclusions are too simplistic: corporate profit alone does not signify a functioning market. 'More market' seems a poor answer to the current conundrum because it may solidify market failure instead of enhancing efficiency (Furubotn and Pejovich, 1972: 1141). Little optimism remains that the online information industry can manage to self-regulate. Currently, there are more reasons to believe that the uninhibited corporate pursuit of online personal data will continue. Below, I discuss some of the institutional frameworks that facilitated today's simultaneous data exchange and extraction on the Internet.

Technocratic fail

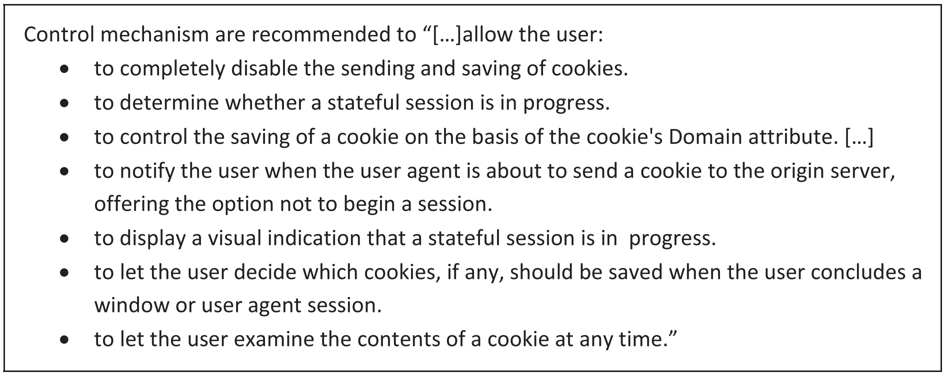
The commercial introduction of the Internet laid the groundwork in the early 1960s for today's many-to-many communication. Originally, the Arpanet installation, the early 1960s predecessor of the Internet, did two things: it opened up secure communication channels that facilitated the collaboration of scientists who worked in far-flung places, and it increased the speed and efficiency of information exchanges and the connectivity (Dumas and Schwartz, 2009). To date, this is still what the Internet does best, but arguably not as securely as originally envisioned. With the introduction of browser cookies, secure online communication has deteriorated.

The history of the browser cookie is briefer than that of the Internet, although in its repercussions for online agency a lot more influential. Before the successful implementation of browser cookies, data exchanges between an Internet user's computer and a remote server were anonymous. That changed in 1994, when Lou Montulli assembled a piece of code in hypertext transfer protocol language (HTML), usually called the HTTP cookie. It was initially labelled a *state information* in the 1995 patent filings and counters memory loss in data exchanges.[11] Four years later in 1998, Netscape, where Montulli was employed, was granted patent number US5774670 A and his invention was awkwardly called a *persistent client state in a hypertext transfer protocol based client–server system* (see patents.com). The patent filings of Montulli's invention describe a way of storing information on the user's computer about a transaction between a user and a server that can be retrieved at a later date by the server.

The persistent and invasive nature of Montulli's invention was recognized immediately by other members of the IETF. An IETF subgroup was initiated to discuss the standards for this budding new technology. Following the most important discussion group threads, it is interesting to observe how engineers were swayed to embrace a piece of code with obvious security shortfalls and invasive protocol transfer properties. The invasive and risky properties of HTTP cookies are recognized but not central in group discussions. Lou Montulli is part of a loosely formed group led by David M Kristol, then head of Bell research laboratories. Regardless of the group's formal openness, it seems the publication of the discussions on this central bit of technology is kept to a select, small, semi-private circle of interested individuals.

One of the initial papers authored by Kristol and Montulli, called the *HTTP State Management Draft*, is a first attempt to outline what is thought to be a group consensus.[12] An extensive section is devoted to privacy problems and discussions on the rights of users to remove or generally cap cookies. Note that these suggestions, if they had been implemented, might have resulted in improved and more secure online data exchanges. Full agency is attributed to online users. The list contained in Figure 2 underscored how far the IETF moved away in its later publications from an original intent to improve online security.

Despite some in-group opposition to the invasive and persistent nature of HTTP cookies, the aforementioned section in the Kristol and Montulli draft of 1997 was removed to 'facilitate convergence' a few weeks later.[13] What is more, the exclusion of user rights to remove and cap cookies turns into a key strategy in the first RFC published (RFC 2109) on the website of the IETF (Kristol and Montulli, 1997).[14] Neither the first nor the second RFC offers more than mere weak support for online agency (Kristol and Montulli, 1997, 2000). The second one matter-of-factly states:



Control mechanism are recommended to "[...]allow the user:

- to completely disable the sending and saving of cookies.
- to determine whether a stateful session is in progress.
- to control the saving of a cookie on the basis of the cookie's Domain attribute. [...]
- to notify the user when the user agent is about to send a cookie to the origin server, offering the option not to begin a session.
- to display a visual indication that a stateful session is in progress.
- to let the user decide which cookies, if any, should be saved when the user concludes a window or user agent session.
- to let the user examine the contents of a cookie at any time."

Figure 2. Working document of the IETF (expired January 1998). *Source*: Internet draft 2.68, 'work-in-progress', pp.16f., kristol.org/cookie/cookie-2.68.txt.

'Informed consent should guide the systems that use cookies' (Kristol and Montulli, 2000: 19). No other references to user rights are incorporated.

The latest update to the *HTTP State Management Draft* dates from April 2011 (Barth, 2011). RFC 6265 contains an insignificant privacy section on HTTP cookies, mentioning a feeble opt-out control mechanism (Barth, 2011). Adam Barth expresses the starkest warnings yet about challenges to Internet users' security and sees users exposed to serious vulnerabilities caused by HTTP cookies (2011: Sect. 8). He emphasises that most user defences are 'ineffective' and outlines how, for example, collaborating servers can inject identifying information into dynamic URLs without even using cookie scripts. Users are 'vulnerable to attacks', 'exposed to eavesdroppers' and 'third party alterations' of their communication, while suffering a 'loss of confidentiality or integrity', 'message interception and redirection' with the possibility of 'gaining the user's authority and confidential information' (Barth, 2011: 30f). Here the consequences of cookie technology are most bluntly stated: the loss of security and machine integrity, personal data loss and user control corruption. An endorsement of any HTTP cookie policy is missing, however.

As a balancing act, the advisory role of the IETF is maintained, while their proposed – and widely ignored – Internet standards accommodate HTTP cookie technology. Because their standards are widely ignored anyway, one might argue that the impact of their accommodating institutional framework is minute. With perfect hindsight, a more pointed statement of resistance and a wider societal debate on this important topic might have been preferable.

Analytically, the *laissez-faire* approach of the IETF exemplifies a case of governmentality, where an institution does not want to intrude on the 'course of things' (Gordon, 1991: 17).[15] Their lack of intervention, public engagement and public debate mediate power by reinforcing the tacit effects of a possible 'natural course' of action (Murray, 2007: 167). Thus, experts in the IETF subgroup seem compelled to accommodate the current development as unavoidable and without alternatives. With only a diffuse public understanding of what web tracking means, a small IETF subgroup spearheads the public decision-making process for the implementation of HTTP cookies. If there is little public input and fewer public policies to influence decision-making in expert institutions, it may seem like a sound course of action to take the path of least resistance within the institution.

HTTP cookies are the most common tracking technology employed on the Internet, but more insidious technologies have been developed over the past years. In itself this development represents another current conundrum for the IETF: their work rapidly seems outdated as new intrusive tracking technologies replace the old ones. Currently, the development of enhanced functionality with new coding styles and script-based web pages spreads vulnerabilities that circumvent user controls (Acar et al., 2014; W3C, 2009). Online users are dealing with embedded objects called 'supercookies', 'zombiecookies', 'uebercookies' or 'evercookies' and these tags are no exaggeration. Circumventing supercookies is almost impossible, given that much of the web content includes videos requiring widely used applications like Adobe flashplayer with in-built backdoors. An online security company offers a stark description:

> And the next generation of supercookies – the Evercookie – takes tracking to even greater heights. The Evercookie can use a multitude of mechanisms to store user data in order to compile unique identities across domains. These mechanisms can include standard tracking cookies, supercookies, Silverlight-isolated storage, RGB values in PNG files, ETags and HTML 5 sessions. (Sheldon, 2013)

A good number of online users would have to look up most of these rarely used terms. ‘Evercookies’ are browser-independent and stored in folders not read by the users’ browsers; they continuously track online activity, are independent of the software used, and cannot be deleted (Narayanan, 2010).

New invasive web-tracking mechanisms include browser and canvas fingerprinting that appear to be spreading with no known user interventions to stop them (Acar et al., 2014; Mowery and Shacham, 2012; Munoz-Garcia et al., 2012). Nobody seems quite sure whether additional tracking techniques exist. Rather more certain, though, is that informed consent of online users is not sought, despite numerous do-not-track initiatives in North America and recent European online consent forms, presented to web users for the placement of HTTP cookies (Dusseault, 2013; Federal Trade Commission, 2012; Lo, 2009; Office of the Privacy Commissioner of Canada, 2012; van Eijk et al., 2012; Vincent, 2012; Whitman, 2004).

An appropriate starting point would be a user agreement that explains invasive technologies or extraction techniques to people who are affected (Pollach, 2007). Further suggestions include behavioural targeting without the use of tracking technologies, a voluntary participation in web tracking for a subgroup of users who choose to be tracked, a pay-per-track app, or the introduction of multiple online profiles with different web-tracking agreements (Eckersley, 2010; Lahlou, 2008; Leber, 2012).

While it is commendable that the IETF issues statements which decry the increasing vulnerabilities on people’s computers caused by web tracking, their emphatic institutional support for a repeal of invasive web-tracking technologies is still outstanding (see, for example, Yen et al., 2012). Its sister organisation, the Internet Society, promotes itself as ‘the world’s trusted independent source of leadership for Internet policy, technology standards, and future development’ (see internetsociety.org), but neither has issued a public statement on the significance of commercially extracted personal data to Internet users.

What has been publicly issued, in a collaborative effort by more than 40 international privacy and security experts, is an extended final version of the ‘International Principles on the Application of Human Rights to Communications Surveillance’.[16] However, commercial web tracking is not mentioned while corporate responsibility is referenced only once:

> Business enterprises bear responsibility for respecting individual privacy and other human rights, particularly given the key role they play in designing, developing, and disseminating technologies; enabling and providing communications; and in facilitating certain State surveillance activities. Nevertheless, these principles articulate the duties and obligations of States when engaging in Communications Surveillance.

The above quote underlines that apparently not only the IETF is mired in what I would describe as ‘passive abstentionism’ (Gordon, 1991: 17). Clearly, state surveillance laws affect the well-being of their citizens, but a relentless submission of online users to their forced data extraction by unseen corporate actors ought to attract equal attention. In the last section, I will show how market failure and technocratic failure explain and solidify the murky situation of online users when they use web search or social media.

Conclusion

Companies that use big social data in combination with web-tracking technologies store personal data on an unprecedented scale. In an unregulated information market and a *laissez-faire* institutional context, the sheer scale of these operations has introduced new challenges to online user agency. Thus, what is the extent of self-regulation that an online user may expect from an online information industry involved in web-tracking technologies? And how does web tracking affect online user agency?

Online corporations operating in a dysfunctional information market do not self-regulate because it puts them at an economic disadvantage, as has become sufficiently clear. Information property rights researchers sometimes advocate for micro payments to reimburse people for the use of their personal data (e.g. Furubotn and Pejovich, 1972; Lanier, 2013; Leber, 2012). This approach misses the point. It fails to understand that ‘[. . .] preserving information is changing from a technological puzzle into a moral dilemma’ (Aiden and Baptiste, 2013: 203). Market liberalization cannot solve this dilemma.

The current unequal online exchange is carried out neither between two fully informed agents nor does it improve the online public sphere. It may be covered, though, by another term that is far less flattering than that of a trade-off: It has the characteristics of an *unconscionable contract*. Unconscionable contracts are exploitative, unjust, unavoidable and put the burden of an economic transaction wholly on one side, and in this case the online user. For most users, the inherent unjust and unequal burden is nowhere spelled out, and even if it were, people would be unlikely to read the terms. In a more humorous take on this serious topic, the Huffington Post reported in 2010 that approximately 7500 online users had ‘accidentally sold their souls’ to an online gaming store (Smith, 2010). Although an absurd example, it underscores how time and again online users are not free to negotiate online contracts; they simply must agree. Online agents are usually not made aware of any deceitful personal data extractions, but even if they were, no meaningful alternatives exist to dodge them – other than going offline. Recent studies on web tracking show that fewer and fewer individual solutions exist to minimise web tracking (Acar et al., 2014; Munoz-Garcia et al., 2012; Sheldon, 2013; Soltani et al., 2009; W3C, 2009; Yen et al., 2012). So with nowhere to turn, online users invariably 'consent' to an unavoidable situation.

Current incentives set by non-regulation nudge corporate actors to engage in more intense web tracking. All the while, Big Data storage capacities are growing and becoming less expensive. What is more, the current under-regulation of online personal data extractions is beneficial for governmental agencies. Numerous incidences exist where, in the past, close connections between the government and information industries were deemed useful (Hedrick, 1991). So, hedged with some caveats, the current wilful political neglect to limit personal data hoarding may be linked to a governmental reliance on increased commercial efforts to extract and store personal data. Any way we look at it, a strong case can be made to direct the scope of the research away from state collectors of personal data towards the unregulated collection of personal data by corporate actors.

Further research ought to look at who is the most heavily targeted population segment that is tracked around the web. Also of interest is the question whether web-tracking companies may be free-riding on the Internet commons, extracting more than their fair share of the available profits. Furthermore, the next emerging path dependency in advancing technology seems to be a convergence on storage instead of computing speed (Shroff, 2013). Part of this trend might be connected to web tracking, indicating a shift away from progressive towards more regressive technological developments.

Declaration of conflicting interests

The author declares that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. Consider the following example of non-regulation: Worldwide, no government seems to have enacted any strict time limit on personal data storage by corporations (personal email, Sam Smith, blogger and technology specialist at privacy international, see his blog entry at https://www.privacyinternational.org/blog/the-cookie-law-is-a-privacy-trainwreck). Some governments allow the use of personal data as long as the corporation has any use for it, but theoretically that covers eternal data storage.
2. I will assume that most online users switch between offline and online social networks, offering a semi-permeable information diffusion between these two communication modes. This ties in with a dominant belief that the World Wide Web is a virtually contained public sphere, where communication is instant and users roam free – usually after paying for hardware, software and access.
3. This user number is a conservative estimate from 2012, see http://www.statisticbrain.com/skype-statistics/
4. In 2011, Microsoft bought Skype for $8.5 billion (see r.reuters.com/vev49r). As a side effect, their acquisition enables them to secure tracked personal data from a host of mobile communication devices.
5. Note that the actual volume of tracked data is unknown because the data points cited here are just the ones stated in their user agreement.
6. Sometimes the courts fill the void of regulatory inactivity. Due to a verdict by the Grand Chamber of the European Court of Justice in May 2014, Google now offers European citizens an online application to erase links, enforcing what is sometimes called the 'right to be forgotten' (see ECJ Case C-131/12, Google Inc. vs. AEPD and Mario Costeja González). Note, though, this court case covers highly visible link entries, not invisible web tracking, discussed below.
7. Public goods are often plagued by the so-called 'free-rider problems' whenever wrong governance systems lead to a deterioration of the commons (Ostrom, 2012). In a way, web-tracking companies are an example of a free-rider problem. They unilaterally and surreptitiously extract personal data, heavily relying on the contributions of all other Internet participants. But this relationship is not at the heart of the current study and left to future analyses.
8. A recent experiment on the manipulative powers of Facebook feeds included approximately 700,000 unaware social media users. Facebook users were shown not all but select feeds, and those with more negative feeds responded more negatively (Kramer et al., 2014). Usually, informed consent is mandatory for human participants in research, but in the social media industry no such requirements exist. Facebook issued a statement saying the study aimed to improve their services, but did not admit any wrongdoing.
9. In an interesting conversation with upper-level students, someone proudly pointed out that they stopped using Facebook when they became aware of the personal data abuse on this platform and now only were using 'whatsapp' to stay in touch with friends. He was unaware that the latter is an affiliate of Facebook and his personal data still were being combined across platforms by the same corporation he had tried to avoid.

10. One may argue that nobody needs to use social media, but that misses the point. For social media users any substitutions for social media are scarce to the point of non-existant.
11. Memory loss means the remote server has no information about the computer requesting the information.
12. This is an early draft, published in July 1997, available at http://kristol.org/cookie/cookie-2.68.txt.
13. See discussion thread 'cookie draft available', available at http://lists.w3.org/Archives/Public/ietf-http-wg-old/1996JanApr/
14. RFC is an acronym for a Request for Comment.
15. I wish to thank an anonymous referee for pointing out this analytical connection.
16. Available at: https://en.necessaryandproportionate.org.

References

Acar G, Eubank C, Englehardt S, et al. (2014) The web never forgets: Persistent tracking mechanisms in the wild. Draft – July 24. Available at: https://securehomes.esat.kuleuven.be/~gacar/sticky/the_web_never_forgets.pdf (accessed July 2014).

Aiden E and Baptiste MJ (2013) *Uncharted: Big Data as a Lens on Human Culture*. New York, NY: Riverhead Books.

Babe RE (1983) Information industries and economic analysis: Policy-makers beware. In: OH Gandy Jr, P Espinosa and JA Ordover (eds) *Proceedings from the tenth annual telecommunications policy research conference*. Norwood, NJ, Annapolis, Maryland: Ablex Publishing Corp., 25–28 April 1982, pp. 123–135.

Baker CE (2002) *Media, Markets, and Democracy*. New York, NY: Cambridge University Press.

Bamford J (1982) *The Puzzle Palace: Inside the National Security Agency, America's Most Secret Intelligence Organization*. Boston: Houghton Mifflin.

Barth A (2011) HTTP state management mechanism, RFC 6265. Available at: http://www.rfc-editor.org/rfc/rfc6265.txt (accessed July 2014).

Benello CG (1981) Technology and power: Technique as a mode of understanding modernity. In: Christians CG and van Hook JM (eds) *Jacques Ellul: Interpretive Essays*. Urbana: University of Illinois Press, pp. 91–107.

Boston Consulting Group (2012) The value of our identity, liberty global policy series. Available at: https://www.bcgperspectives.com/content/articles/digital_economy_consumer_insight_value_of_our_digital_identity/ (accessed September 2014).

Bourdieu P (1990) *The Logic of Practice*. Stanford, CA: Stanford University Press.

Chang RM, Kauffman RJ and Kwon YOK (2014) Understanding the paradigm shift to computational social science in the presence of Big Data. *Decision Support Systems* 63: 67–80.

Chen M, Shiwen M and Yunhao L (2014) Big Data: A survey. *Mobile Network Applications* 19: 171–209.

Crump C and Harwood M (2014) Invasion of the data snatchers: Big Data and the internet of things. *ACLU Blog,* 25 March. Available at: https://www.aclu.org/blog (accessed November 2014).

Cumbley R and Church P (2014) Is "Big Data" creepy? *Computer Law and Security Review* 29: 601–609.

Curran J (1991) Mass media and democracy: A reappraisal. In: Curran J and Gurevitch M (eds) *Mass Media and Society*. London: E. Arnold, pp. 82–117.

Denardis L (2014) *The Global War for Internet Governance*. New Haven, CT: Yale University Press.

Doctorow C (2008) Welcome to the petacenter. *Nature* 455: 16–21.

Dwoskin E (2014) In a single tweet, as many pieces of metadata as there are characters. *Digits (Blog),* 6 June. Available at: http://blogs.wsj.com/digits/2014/06/06/in-a-single-tweet-as-many-pieces-of-metadata-as-there-are-characters/ (accessed November 2014).

Dumas MB (2012) *Diving into the Bitstream: Information Technology meets Society in a Digital World*. New York, NY: Routledge.

Dumas MB and Schwartz LM (2009) *Principles of Computer Networks and Communications*. Upper Saddle River, NJ: Pearson Prentice Hall.

Dusseault PL (2013) Report on the standing committee on access to information, privacy, and ethics. Fifth report. Available at: http://www.parl.gc.ca (accessed October 2013).

Eckersley P (2010) How unique is your web browser? In: Mikhail JA and Nicholas JH (eds) *Privacy Enhancing Technologies*. Berlin: Springer, pp. 1–18.

El Akkad O (2014) The strange connection between the NSA and an Ontario Tech Firm. *The Globe and Mail*, 20 January. Available at: http://www.theglobeandmail.com/technology/business-technology/the-strange-connection-between-the-nsa-and-an-ontario-tech-firm/article16402341/ (accessed July 2014).

Federal Trade Commission (2012) Protecting consumer privacy in an era of rapid change. Recommendations for businesses and policy makers. Available at: http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf (accessed September 2014).

Furubotn EG and Pejovich S (1972) Property rights and economic theory: A survey of recent literature. *Journal of Economic Literature* 10(4): 1137–1162.

Gandy OH Jr (1993) *The Panoptik Sort: A Political Economy of Personal Information*. Boulder, CO: Westview Press.

Gordon C (1991) Governmental rationality: An introduction. In: Burchell G, Gordon C and Miller P (eds) *The Foucault Effect: Studies in Govermentality*. Chicago: University of Chicago Press, pp. 1–52.

Habermas J (1989) *The Structural Change of the Public Sphere*. Cambridge, MA: MIT Press.

Hedrick D (1991) *The Invisible Weapon*. New York: Oxford University Press.

ITU (2014) *The World in 2014 - ICT Facts and Figures*. Geneva, Switzerland: International Telecommunication Union Data and Statistics Division. Available at: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2014-e.pdf (accessed June 2013).

Jussawalla M and Cheah CW (1987) *The Calculus of International Communications*. Littleton, CO: Libraries Unlimited.

Kambatla K, Kollias G, Kumar V, et al. (2014) Trends in Big Data analytics. *Journal of Parallel and Distributed Computing* 74: 2561–2573.

Knight A and Saxby S (2014) Identity crisis: Global challenges of identity protection in a networked world. *Computer Law and Security Review* 30: 617–632.

Korupp SE (2006) No man is an island: The influence of knowledge, household settings, and social context on private computer use. *International Journal of Internet Science* 1(1): 45–57.

Korupp SE, Kühnemund H and Schupp J (2006) Die digitale Spaltung in Deutschland. Geringere Bildung – seltener am PC. *German Institute for Economic Research in Berlin Weekly Report* 19(6): 288–294.

Kramer ADI, Guillory JE and Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111(24): 8788–8790.

Krishnamurthy B, Malandrino D and Wills CE (2007) Measuring privacy loss and the impact of privacy protection in web browsing. In: *Symposium on usable privacy and security (SOUPS)*, Pittsburgh, PA, 18–20 July. Available at: http://web.cs.wpi.edu/~cew/papers/ (accessed July 2014).

Kristol DM and Montulli L (1997) HTTP state management mechanism, RFC 2109. Available at: http://rfc2109.openrfc.org/ (accessed July 2014).

Kristol DM and Montulli L (2000) HTTP state management mechanism, RFC 2965. Available at: http://rfc2965.openrfc.org/ (accessed July 2014).

Lahlou S (2008) Identity, social status, privacy, and face-keeping in digital society. *Social Science Information/Information Sur Les Science Sociales* 47(3): 299–330.

Lanier J (2013) *Who Owns the Future?* New York, NY: Simon & Schuster.

Laredo Group Inc. (n.d.) Laredo group glossary. Available at: http://www.laredogroup.com/free-tools-resources/index.asp (accessed July 2014).

Leber J (2012) A dollar for your data. *MIT Technology Review - Business Report*. Available at: http://www.technologyreview.com/news/428046/a-dollar-for-your-data/ (accessed November 2014).

Lo J (2009) A "Do Not Track List" for Canada? *Public Interest Advocacy Center (PIAC) Report*. Available at: http://www.piac.ca/files/dntl_final_website.pdf (accessed June 2013).

Lu R, Zhu H, Liu X, et al. (2014) Toward efficient and privacy-preserving computing in Big Data era. *IEEE Network* July/August: 46–50.

Minister of Industry (2010) Internet shopping in Canada: An examination of data, trends and patterns. Available at: http://www.statcan.gc.ca/pub/88f0006x/88f0006x2009005-eng.pdf (accessed June 2013).

Mowery K and Shacham H (2012) Pixel perfect: Fingerprinting canvas in HTML5. In: *Proceedings of W2SP 2012*. San Francisco, USA: IEEE Computer Society, May 2012. Available at: http://w2spconf.com/2012/papers/w2sp12-final4.pdf (accessed November 2014).

Munoz-Garcia O, Monterrubio-Martin J and Garcia-Aubert D (2012) Detecting browser fingerprint evolution for identifying unique users. *International Journal of Electronic Business* 10(2): 120–141.

Murray KB (2007) Governmentality and the shifting winds of policy studies. In: Orsim M and Smith M (eds) *Critical Policy Studies*. Vancouver: UBC Press, pp. 161–184.

Narayanan A (2010) Cookies, supercookies and ubercookies: stealing the identity of web visitors. *33 bits of Entropy Blog,* 18 February. Available at: http://33bits.org/2010/02/18/cookies-supercookies-and-ubercookies-stealing-the-identity-of-web-visitors/ (accessed November 2014).

Numan D and DiDomenico ML (2012) Market research and the ethics of Big Data. *International Journal of Market Research* 55(4): 2–13.

Office of the Privacy Commissioner of Canada (2012) *Report on the 2010 Office of the Privacy Commissioner of Canada's consultations on online tracking, profiling and targeting, and cloud computing*. Available at: https://www.priv.gc.ca/resource/consultations/report_201105_e.pdf (accessed September 2014).

O'Harrow R Jr (2005) *No Place to Hide*. New York, NY: Free Press.

Ostrom E (2012) The future of the commons: Beyond market failure and government regulations. In: Ostrom E, Chang C, Pennington M, et al. (eds) *The Future of the Commons*. London: Institute of Economic Affairs, pp. 68–83.

Parreira AT (2013) Methods and systems for real-time web-tracking and marketing. Patent application US 2013/010403 A1, USA.

Peacock SE (2014) We are the data! Big data and user agency. In: *Social media and society, September 26/7 (Poster Presentation)*, Ted Rogers School of Management, Ryerson University, Toronto.

Peacock SE and Kühnemund H (2007) Senior citizens and young technologies: Reasons for senior citizens' non-access and access of the internet in a European comparative perspective. *European Journal of Aging* 4(4): 191–200.

Pollach I (2007) What's wrong with online privacy policies? *Communications of the ACM* 50(9): 103–108.

Roderick L (2014) Discipline and power in the digital age: The case of the US consumer data broker industry. *Critical Sociology* 40(5): 729–746.

Sheldon R (2013) Supercookies take a bite out of enterprise desktop security. Available at: http://searchenterprisedesktop.techtarget.com/tip/Supercookies-take-a-bite-out-of-enterprise-desktop-security (accessed June 2013).

Shroff G (2013) *The Intelligent Web: Search, Smart Algorithms, and Big Data*. Oxford: Oxford University Press.

Smith C (2010) 7,500 online shoppers accidentally sold their souls to gamestation. *Huffington Post*, 17 June. Available at: http://www.huffingtonpost.com/2010/04/17/gamestation-grabs-souls-o_n_541549.html (accessed November 2014).

Soltani A, Canty S, Mayo Q, et al. (2009) Flash cookies and privacy. SSRN manuscript. Published 10 August 2009.

Available at: http://ssrn.com/abstract=1446862 or http://dx.doi.org/10.2139/ssrn.1446862 (accessed September 2013).

Statistics Canada (2011) *Canadian Internet Use Survey 2010*. Machine readable data file.

Statistics Canada (2013) *Canadian Internet Use Survey 2012*. Machine readable data file.

Turow J (2012) The disconnect about what people say and do about privacy. *Journal of Law* 2: 479–482.

van Eijk N, Helberger N, Kool L, et al. (2012) Online tracking: Questioning the power of informed consent. *Info* 14(5): 57–73.

Vincent N (2012) What is do not track? *FTC Consumer Education Specialist Blog*. Available at: http://www.onguardonline.gov/blog/what-do-not-track (accessed June 2013).

W3C (2009) Document Object Model (DOM). Available at: http://www.w3.org/DOM/ (accessed July 2014).

Weide K (2011) Real-time bidding in the United States and Western Europe, 2010-2015. *IDC White Paper*. Available at: http://info.pubmatic.com/rs/pubmatic/images/IDC_Real-Time%20Bidding_US_Western%20Europe_Oct2011.pdf (accessed September 2014).

Whitman JQ (2004) The two western cultures of liberty: Dignity versus liberty. *The Yale Law Journal* 113(6): 1151–1221.

Wondracek G, Holz T, Platzer C, et al. (2010) Is the Internet for porn? An insight into the adult online industry. In: *9th workshop on the economics of information security (WEIS)*, USA, June 2010. Available at: https://www.cs.ucsb.edu/~chris/research/doc/weis10_pron.pdf (accessed September 2014).

World Economic Forum (2011) Personal data: The emergence of a new asset class. Available at: http://www.weforum.org/reports/personal-data-emergence-new-asset-class (accessed May 2014).

Yen TF, Xie Y, Yu F, et al. (2012) Host fingerprinting and tracking on the web: privacy and security implications. Internet Society Report. In: *NDSS Symposium 2012*, San Diego, USA. Available at: http://www.internetsociety.org/sites/default/files/11_3.pdf (accessed September 2014).