NEW ECONOMICS FOUNDATION

THE RISE OF THE DATA OLIGARCHS

POWER AND ACCOUNTABILITY IN THE DIGITAL ECONOMY

PART 1: DATA COLLECTION

Written by: Duncan McCann

New Economics Foundation
www.neweconomics.org
info@neweconomics.org
+44 (0)20 7820 6300
@NEF

CONTENTS

PREFACE

A new economy is emerging. And this new economy is powered by a new type of fuel: data. As the data economy becomes increasingly prominent, there are troubling signs that it is worsening existing power imbalances, and creating new problems of domination and lack of accountability. But it would be wrong simply to draw dystopian visions from our current situation. Technological change does not determine social change, and there is a whole range of potential futures – both emancipatory and discriminatory – open to us. We must decide for ourselves which one we want.

This is the first of four papers exploring power and accountability in the data economy. These will set the stage for future interventions to ensure power becomes more evenly distributed. This paper explores the impact of the mass collection of data, while future papers will examine: the impact of algorithms as they process the data; the companies built on data that mediate our interaction with the digital world; and the labour market dynamics that they are disrupting.

Our research so far has identified a range of overarching themes around how power and accountability is changing as a result of the rise of the digital economy. These can be summarised into four arguments:

- **Although the broader digital economy has both concentrated and dispersed power, data has had very much a concentrating force.**
- **A mutually reinforcing government-corporation surveillance architecture – or data panopticon – is being built, that seeks to capture every data trail that we create.**
- **We are over-collecting and under-protecting data.**
- **The data economy is changing our approach to accountability from one based on direct causation to one based on correlation, with profound moral and political consequences**

This four-part series explores these areas by reviewing the existing literature and conducting interviews with respected experts from around the world.

INTRODUCTION

The Facebook/Cambridge Analytica scandal has made data gathering a front-page story in recent months. We have identified four key issues related to data gathering:

- **GDPR will not save us:** Although the General Data Protection Regulation (GDPR) will be an improvement for data privacy, it should not be considered a panacea. Some companies, especially global ones, will structure their business to dodge the regulations.
- **Privacy could become the preserve of the rich:** The corporate data gathering industry may evolve to create a system where only the rich are able to afford the necessary tools and labour time to effectively maintain their privacy.
- **Privacy is an increasingly unmanageable burden:** responsibility for managing data falls far too heavily on the individual rather than those who want to use individuals' data.
- **Are we becoming a conformist society?** Ubiquitous data collection, coupled with data never being deleted means we could be entering an era of self-censorship and 'social cooling'.

1. BACKGROUND

Data sustains the modern digital ecosystem, from online services, to apps, to websites. Were data not being gathered during every digital interaction, and then processed and monetised, today's digital economy would be very different. Would society have paid for the nascent services we now rely on on a daily basis? Would we have free access to online maps, translation services, email programmes, and other services? Almost certainly not.

Long before even the Industrial Revolution, in 1597 Sir Francis Bacon wrote in *Meditationes Sacrae* that "knowledge itself is power". The novelist Tom Clancy rephrased it to say that "[i]nformation, knowledge, is power. If you control information, you can control people". These two quotes offer a tantalising insight into one of the key issues with the digital economy's reliance on the ubiquitous gathering of data: namely, that we are conferring huge amounts of power to those entities that are best able to gather and process this data.

The ubiquitous nature of data collection today is shifting the very notion of privacy. Indeed, some have claimed that we are living in a post-privacy world.[1] For others, the debate about privacy centres on the need to ensure "individual control over information flows."[2] This is a limiting view which obscures "how and why powerful institutions use data to nudge us toward their own economic and political ends."[3] The New York Times bestseller, *Dragnet Nation* by Julia Angwin, articulates this idea: that an individual's privacy has everything to do with power and the prospect of manipulation, mostly hidden from view and without their knowledge and understanding.[4]

Orwell's conception of 'Big Brother', from the dystopian novel *1984,* looms large in debates over privacy. Much work has focused on preventing loss of privacy, and its secondary effects of self-censorship, embarrassment, or loss of reputation. Regulation and other remedies have focused on trying to protect us from this. However the nature of ubiquitous data collection, together with the mass processing of data by thousands of different companies and algorithms, means that concerns around this Orwellian vision are only half the story. While those dangers are still present, a new world is emerging which is characterised by "a thoughtless process of bureaucratic indifference, arbitrary errors, and dehumanization, a world where people feel powerless and vulnerable, without any meaningful form of participation in the collection and use of their information."[5]

The relationship between our online life and our attitude to privacy is complex. As a society we are more aware than ever of privacy issues, through the frequent news stories

of embarrassing data leaks and the growing digital dossiers about us held by data brokers. Our actions however often do not mirror these concerns.[6] Today, we are sharing more personal data than ever before, without being given the tools to secure our privacy.[7] Today, while we search for information or watch entertainment, the companies behind these sites are busy monitoring and gathering data about us.[8]

At the core of this issue of gathering data and privacy is a transparency paradox: "Big data promises to use this data to make the world more transparent, but its collection is invisible, and its tools and techniques are opaque, shrouded by layers of physical, legal, and technical privacy by design. If big data spells the end of privacy, then why is the big data revolution occurring mostly in secret?"[9]

Big data is already causing a shift in the distribution of power in the economy. Although in general the digital economy both concentrates and disperses power, big data seems to amplify the concentration effect while minimising the dispersal effect. This is because big data is likely to "benefit the institutions who wield its tools over the individuals being mined, analysed and sorted."[10] And, since these tools are predominantly, in the hands of large companies and government agencies, most people will be excluded from the empowerment that such tools allow. The exceptions are those people who are able to understand and use the tools, and are thus able to expose truth as to power, as in the case of the *Who Owns England* blog, which uses data gathering and analysis to make visible something which is normally invisible – in this case, who owns the land in England.[11]

1.1 A HISTORY OF DATA GATHERING

In the world of 'small data' (which was not so long ago) data collection looked very different. Prior to the collection of data, a detailed analysis was done to understand exactly which data points were needed for the specific purpose of the collection. Since there was significant cost to collection and verifying each data point, at every stage people would ensure that the data collected was limited to only the amount that necessary for the job at hand, Data storage was also expensive, meaning that once the data had been used, it was often deleted, unless the company could foresee an ongoing purpose for it.

In the early days of the Internet it was not easy to track individuals across multiple platforms without requiring them to repeatedly login and authenticate themselves. This all changed with the invention of the cookie, and the Internet went from "being a relatively anonymous activity, like wondering the streets of a large city, to the kind of environment where records of one's transactions, movements and even desires could be

stored, sorted, mined and sold."[12] Website owners and data companies use trackers and cookies embedded in users' browsers to literally follow them around the Internet, recording as much as they can.

These cookies are installed as we surf the net and are used by website owners but also by third parties wanting to gather data on individuals. A detailed study by the Wall Street Journal found that, of the 50 most popular websites, only Wikipedia did not transfer data to third parties. Almost half transferred data to 60 third parties, and the worst offender, dictionary.com, shared data with 234 third parties.[13] Although researchers found that individuals' regular use of prominent companies like Google, Twitter and Facebook were nearly ubiquitous, much of the data sharing was occurring with companies who remain largely hidden from public view.[14]

There is a growing understanding that the products that we use are tracking us. Our browser and the websites we visit are recording all of our online activity, and the smartphone in our pocket and the apps we download are sharing usage and location data. These companies then use the data gathered within their domains to improve their services and make more revenue, either through additional sales or through selling raw or processed data to other companies.

A new frontier is now opening up with the increase in popularity of fitness trackers, driving monitoring devices, and the 'internet of things', all of which are generating vast amounts of data for collection.[15]

1.2 KEY TRENDS

Data has become so important to the digital economy that for many companies, data gathering is fundamental to their existence, and changes in technology or regulation could upset their business models. For instance, the recent inclusion of robust tracker blocking technology in Apple's browser Safari will prevent users being 'followed around' the Internet.[16] Other browsers like Firefox and Google Chrome have decided to follow suit.[17] Although this may be broadly welcomed by browser users, the data gathering industry is getting worried. The digital advertising technology company Criteo predicts that tracker blocking in Safari, as well as closing some loopholes in the mobile iOS version of the browser, will mean a 22% reduction in revenues for the coming year.[18] And as well as changes to browsers, companies are developing add-ons to help people manage their data trails and restrict the amount of data that can be gathered about them while they surf.[19]

The amount of data being gathered is growing exponentially, with 90% of the world's data created in the last 2 years.[20] People are spending more and more time online (20 hours a week for the average person in the UK,[21] and over four hours per day on a smartphone for people in the USA[22]) and more parts of their lives are mediated by digital technology and networks.[23] We are sharing ever-increasing amounts of data, some of it extremely sensitive, like information about our dating preferences, health, and finances.

In the past, gathering, storing and processing data was hugely expensive and so most entities gathered the bare minimum. Advances in technology have changed this entire sector. Data gathering is now cheap and easy. This means that companies are not just gathering the necessary data and using it right away, but also storing it in the hope that they will be able to unlock future value from it.[24]

There are two primary reasons why entities gather data. Firstly, they want to improve the quality of their data refineries, or algorithms, and the accuracy of their outcomes. Most algorithms improve with the amount of real world data one is able to feed into them. A simple example would be Amazon's recommendation algorithm: the more data that Amazon is able to collect (about users' purchases, preferences and site browsing habits) the better it is able to generate correlations that may be useful to other users – but with the ultimate purpose of driving more sales within Amazon. Secondly, processed data is used to target adverts at users. The principle is that the adverts can be personalised so that people only see adverts that are relevant to their general interests or respond to specific online activity. The Holy Grail for these new digital advertising agencies is to stop adverts being a nuisance, and for them to be considered useful and non-intrusive.

2. ISSUES

2.1 THE DECEPTIVE MEANING AND USE OF 'CONSENT'

Consent is one of the most discussed issues with regard to collecting and using data gathered from our digital activity. The existing UK Data Protection Act 1998 (DPA) states that one of the conditions for a companies to gather and process 'personal data'[25] is to have obtained consent from the individual for the specific data to be collected and processed in a specific way or for a specific purpose.[26] Although the Act does not define 'consent', the EU directive from which it emanates defines it as "any freely given specific and informed indication of his [sic] wishes by which the data subject signifies his agreement to personal data relating to him being processed."

Consent is often used by UK data controllers as the sole legitimising fair processing condition (or sometimes as a back-up to another fair processing condition or grounds for processing) because it is the easiest condition or mechanism by which the data controller can show that they have complied with the DPA 1998. However, according to legal advice:

> *That is not to say that this is always the best condition or ground for data controllers to rely on. In actual fact, it can often be a poor way to secure compliance. This is because individuals may withhold their consent, their consent may be withdrawn … or indeed the reasons for which consent was originally sought and granted may have changed. In the latter case, this would mean that the data controller could no longer rely on the consent originally given.*[27]

Today, it is highly debatable whether people do really provide the type of consent that the original directive conceived of, especially given that the directive was drafted in a pre-Big Data world. This has allowed data gathering to proliferate widely with very little accountability, with ever greater power accumulating in the hands of data warehouses, leaving regulators playing catch-up. In the UK, the Information Commissioner's Office (ICO) has launched an inquiry into the use of data profiles on social media platforms to create targeted political messages.[28] The Belgian government has gone one step further and demanded Facebook stop collecting data on users, or face fines of up to €100 million.[29] Yet whilst these are positive steps, they only scratch the surface of the myriad of ways in which data about us is collected and monetized.

Cate has argued that the original DPA focuses too heavily on the notion of informed consent,[30] which is especially problematic when the empirical evidence shows that

individuals neither read nor understand company data collection and sharing policies.[31] We need to update the rules governing data collection to reflect this reality.

There has been a lot of discussion about the new data protection regime that is being ushered in with the General Data Protection Regulation (GDPR), which will come into force in Europe, including the UK, in May 2018. The GDPR enhances our rights as individuals with regard to consent, and places additional burdens on the data gathering companies. The new definition of 'consent' builds on the original EU directive and requires that consent be unambiguous and involve clear affirmative action. We may now also withdraw consent at any time and companies must make it as easy to withdraw consent as to give it.

The guidance from the ICO summarises the efforts of the GDPR as changes to "reflect a more dynamic idea of consent: consent as an organic, ongoing and actively managed choice, and not simply a one-off compliance box to tick and file away."[32]

But all this focus on consent obfuscates the fact that many companies' use of data is clearly not consented to in any meaningful way, and that much of the data gathered is classified as non-personal so as to escape the reach of DPA and GDPR. And this is all perfectly legal. Indeed the ICO itself has issued a guidance document in order to bust the myth that consent is needed to process personal data.[33] And this has some common sense to it. It would not make sense for banks sharing information about potential criminal transactions, or insurance companies processing claims, to seek consent from all the parties involved. In fact there are six legal bases for processing data – with consent being one, but not accorded any higher status that the alternatives (see Figure 1).

Although the Act provides these six legal bases for collecting data, the evidence is that many data brokers exempt themselves from the whole legislative framework by claiming "to deal with anonymous data, or deny being a data controller, or structure their operations in order to avoid EU jurisdiction."[34] It is hard to understand what impact the GDPR, which only applies in the EU, will have on players within the global data industry, who have a track record of actively seeking to circumvent compliance with data protection legislation. The GDPR does specifically mention fraud prevention, direct marketing and network security as examples of legitimate interest purposes.[35] The GDPR's changes in the nature of granting consent, which must now be active and specific, means that in the future it is likely that fewer and fewer data brokers will rely on consent to justify their data collection.

Figure 1. Lawful basis for processing under GDPR

(a) Consent: the individual has given clear consent for you to process their personal data for a specific purpose.

(b) Contract: the processing is necessary for a contract you have with the individual, or because they have asked you to take specific steps before entering into a contract.

(c) Legal obligation: the processing is necessary for you to comply with the law (not including contractual obligations).

(d) Vital interests: the processing is necessary to protect someone's life.

(e) Public task: the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law.

(f) Legitimate interests: the processing is necessary for your legitimate interests or the legitimate interests of a third party unless there is a good reason to protect the individual's personal data which overrides those legitimate interests. (This cannot apply if you are a public authority processing data to perform your official tasks.)

Figure 1: Article 8 GDPR 2016/679

So despite the collection and resale of personal data for an unknown purpose generally not being permitted within the EU, many data brokers are able to skirt the meat of the legislation and continue to collect and resell data. [36]

The GDPR expressly introduces a legal accountability obligation to European data protection law. While short in length, the new provisions are likely to have far-reaching consequences in practice. Data controllers will be required to establish appropriate technical and organisational measures to ensure that data processing is performed in accordance with the GDPR. They will also be required to be able to demonstrate that the processing is GDPR-compliant when requested by a relevant authority.[37]

The GDPR aims to give individuals more power over the data that is collected about them. Although this is a laudable goal, it is also problematic when analysed in detail. Firstly, it is hard to understand how people can meaningfully consent to sharing data when you consider the number of entities capturing the data, and the variety of purposes for which it is being gathered. It would require people to read long terms and conditions to authorise the data sharing and potentially require addition permission, along with further terms and conditions, to consent to subsequent uses once the data has been sold or merely being used for something that was not in the original contract. We explore this in more detail in our upcoming publication on data processing.

Secondly, people also feel powerless when confronted with decisions about whether to share data. In many cases people are unaware of how much data is being collected about them. People are also unaware about the purpose of the data gathering. Often, even the company gathering the data does not know what it is going to do with it. Without halting the use of big data entirely, it does not make sense for the company in question to try to gain the active consent of all the data subjects for a new piece of data analysis. The market for personal data in many ways relies on customers' lack of awareness of its functioning to keep working.

2.2 THE BATTLE OVER ONLINE PRIVACY

The power of the data collection industry is growing all the time as it collects more and more information about us in order to provide us with relevant ads, filter our job applications, and approve or reject us for credit. To balance this, users have been provided with a suite of tools to enable us to limit what data we share, from the ability to change our device's settings to installing add-on programmes to our browsers.. However, many of us do not use them and it is debateable whether many of the tools offer us any meaningful protection and control.

Only personally identifiable data is protected by legislation, leading many to store and share datasets that have been de-identified or anonymised. Big data techniques challenge the meaningfulness of the legal distinction between personal and non-personal data, as well as the more fundamental notions of privacy, because they enable the re-identification of data subjects using multiple datasets to cross-reference. This renders anonymisation impossible.[38]

Does this mean that we should consign privacy to history? In the arms race which pits data collectors' needs and their economic and technical muscle against individuals' desire for privacy, one side holds most of the power. When you consider the scale of potential leaks, both criminal and negligent (such as the hacking of 3 billion Yahoo accounts)[39] it is easy to think that we have already lost the war.

Personal information that an individual would not want shared publically can surface in a number of ways. Firstly, there are instances where data is shared that exposes private information against the wishes of the data subject, but is in line with the data holder's policies. One example is the now famous story of the young Target customer in the US who was sent vouchers for pregnancy-related items before she had informed her parents of her pregnancy. The information which determined what vouchers she would be sent was purely based on associated purchases.[40] Secondly, there are instances of data released due to negligence on the part of the data holder. Finally, data can be stolen –

although often negligence plays a part in this, as in the case of hackers stealing 145 million Equifax records. This was possibly the most serious data breach in history because of the sensitive personal financial information that was stolen.[41]

We need to find a way to ensure that people can continue to share personal and sensitive information online without the fear of it becoming public.

One big concern in the literature is that we may evolve a system where only the rich are able to afford the necessary tools and manpower to effectively maintain privacy. Mark Zuckerberg provides an interesting illustration of this. He has stated that "privacy is dead"[42] and "no longer a social norm"[43] but, when it comes to his own privacy, has said he will fight hard to protect it. When he bought his house in 2013 he also bought the four adjacent properties and forced all contractors to sign extensive non-disclosure agreements. It is vital that we do not allow privacy to be the exclusive right for the rich.[44]

2.3 THE PROBLEM WITH MAKING INDIVIDUALS RESPONSIBLE

It is generally considered to be the responsibility of the individual to manage their own data trail. Growing disquiet about the amount of data being gathered has resulted in companies providing us with a suite of settings, plugins and consent forms supposedly to empower us to manage our data. However, most people do not understand the pervasive nature of data collection today, do not use the tools at their disposal, or falsely believe that policy and regulation already protects them. Often people only start to engage with their ability to limit the sharing of personal data once they have been the subject of a breach or bad experience.[45] The reliance on individual data management is misguided. Developments in car safety point to a way forward: a system which legislates driving conditions and the correct safety features for vehicles, supplemented by relying on personal behaviour only were circumstances deviate from the norm.

A recent paper by Bergstrom found that the "more people trust others, the less concern they have for misuse of personal information"[46] which leads people to be less anxious when online and more inclined to share data. Research by Baek et al. showed that many people lack the knowledge about how to protect their identity and personal information online, with most not engaging in any action to limit the sharing of their data.[47]

In 2006 Barnes contended that "adults are concerned about invasion of privacy, while teens freely give up information …. [and] this occurs because often teens are not aware of the public nature of the Internet."[48] Although this may have been a true reflection of the varying attitudes between generations over a decade ago, more recent research by

Blank et al. in 2014 found that in fact "people who check and change their privacy settings tend to be young."[49] A detailed study of the attitudes of American young people also found that they "have an aspiration for increased privacy even while they participate in an online reality that is optimized to increase their revelation of personal data."[50] Blank et al. go on to describe a "new privacy paradox" which is that "social life is now conducted online and that the social networks do not provide users with the tools that would adequately enable them to manage their privacy settings in a way that is appropriate for them."[51]

There is a growing body of literature that, while recognising that the individual has a significant part to play in controlling the amount of data that they share, argues that industry and government need to do a lot more to ensure that the majority who take no proactive action to protect themselves are at least somewhat protected.[52] [53] [54]

There are cases where we are, in essence, powerless to avoid sharing data, such as sending an email. Some sites, apps and services, such as Facebook, or email are fundamental to being part of a modern society and it can hard to not engage with these providers. This often leaves people sharing data because they have no other option.

An additional complicating factor is that much of the data that is collected about us is gathered from data submitted by others or by inferences made because of our connections to others. It is therefore important to heed Danah Boyd's point that "it's no longer about what you do that will go down in your permanent record. Everything that everyone else does that concerns you, implicates you, or might influence you will go down in your permanent record."[55] This leads to what Niseenbaum and Barocas term "the tyranny of the minority" where a small group, explicitly consenting to share data with a refinery, forces the majority to be subject to the same data collection and analysis, despite having no meaningful or recognized relationship with the consenters.[56] This fundamentally challenges the notion that consent and its removal are meaningful in a world of big data. A study by Mislove et al. showed that when as little as 20% of users reveal attributes, this can be used to infer those attributes in the whole population.[57]

The more systemic response is to acknowledge that people continue to want some degree of privacy and control over the data that they share. We should therefore focus on the reforming the current online environment which forces people to divulge personal information in order to prevent potential social inclusion from not participating in social media platforms.[58]

2.4 DATA PERMANENCE

The old model of small data collection outlined in Section 1.1 is no longer with us. We have entered a world of ubiquitous data gathering. Digital storage has become so cheap that it can make economic sense to store all data collected in perpetuity. This has a big impact on the information held about individuals' past actions. It presents the prospect of people being held accountable for actions in the distant past that may have otherwise been forgotten.

Of course in some ways this can be a positive development, since people may question the rights of serious offenders to escape accountability for their past. We are already seeing the impact of this, with a number of high profile cases of public appointments being made only for their social media history, sometimes over 10 years old, to be resurrected to show them to be unsuitable for the job. One recent case was of Toby Young's appointment to the University regulator, shortly after which where many old tweets resurfaced, which showed that he had held misogynist views about women. He claimed that he had reformed, while others claimed that these showed how unsuitable he was for the post.[59]

The following quote illustrates the problem well:

> *Everything that's on file about you for the last 15 years and the next 40 years may someday be used against you with technology that, at this time, we can't understand or predict. And much of the information that we leave in our wake has no legal protection from being sold in the future: We overcollect and we underprotect.*[60]

One of the responses to the data permanence question has been to bestow individuals with a new power: the 'right to be forgotten', first articulated in EU case law in 2014. The right was an extension of the 1995 Data Protection Protocol which gave individuals and 'authorities' the right to demand the erasure of certain types of material.[61] The 2014 case concerned a Spanish citizen, who complained that an auction notice of his repossessed house continued to appear when searching for his name in Google.[62] Since the matter was completely resolved, he contested that it should no longer be visible as, amongst other things, it could prejudice his ability to get a mortgage.

The European Court of Justice ruled on three key questions. Firstly, that companies would be accountable to the court so long as they had a subsidiary or office in the country. Secondly, that search engines were controllers of data and therefore subject to the regulations on data. And finally, that individuals may also, under certain circumstances, have the right to remove links to information that is "inaccurate, inadequate, irrelevant or excessive" for the purpose of data processing. There is always a

complex balancing act between the rights to freedom of expression, and so cases need to be decided on a case-by-case basis. The EU ruling has inspired other jurisdictions to create their own policy framework. The Canadian Supreme Court declared in 2017 that individuals could require links to be deleted globally from search engines,[63] going further than the EU which merely required them to be deleted within the EU (although the French took a different approach mirroring the Canadians).[64] A Japanese case from 2017 demonstrated how the right of the public to know should trump an individual's right to privacy when stating that a convicted paedophile had not right to have links to his conviction erased.[65] The right to erasure is now enshrined in Article 17 of GDPR.

The right to be forgotten attempts to rebalance power between individuals and data companies by empowering individuals to control what information can be easily accessed about themselves. It has to be balanced against the rights of others to express themselves, but also needs to strike the right balance with regard to holding people to account for their past actions. Case law shows that minor infractions – like having your house repossessed – can be erased, whereas very serious offences – paedophilia – cannot. Exactly where our individual power ends and societal accountability starts is hard to define and will only be clarified through additional case law.

This exposes another serious problem with the right to be forgotten. Since data controllers can be fined for non-compliance, there will be an incentive for data controllers to err on the side of caution and delete borderline cases rather than fight in the court for them to remain online. Tech companies are then acting as appropriate censors of information. In effect, Google, by acting as the gatekeeper of what information can be accessed through their results, is exercising immense power without any meaningful oversight from the state.

If Google declines the removal of information then this opens the way for state involvement, where case law will balance the relevant rights against one another. However, if Google approves the removal of information then that information is gone. There is no obvious way for the public to complain against poor balancing. Google has already gone some way to addressing this problem by contacting the sites from which links are being removed. This at least allows these website owners to represent their or the public's interest.[66] Incredibly, by September 2016, Google had removed over 1.7 billion URLs from the search results, following over 560,000 requests – with only 40 million URL removals rejected.[67] Although they have stopped reporting on the total number of delisted URLs, they do report on the percentage of total removal requests they grant, which currently stands at 43.3%.[68]

2.5 SELF-CENSORSHIP AND SOCIAL COOLING

The confluence of the increase in data collection, the reduction in control and privacy over individuals' data, and of data being stored permanently more often, are ushering in a new world. One of the unforeseen consequences could be what is being called 'social cooling', defined as the self-censoring and pro-conformity implications of ubiquitous, un-private, permanent data collection.[69] The inventor of the term fears for a world characterised by "increasing social conformity and rigidity, in which we self-censor or second guess what we do online for fear of repercussions."[70]

The effects have already been noted when analysing the use of various search terms and websites after the Snowden revelations.[71] A study by Marthews and Tucker found that search terms that were seen as sensitive to issues of privacy saw a decrease in searches and that, perhaps surprisingly, the biggest effects were in countries considered US allies.[72] Another study additionally found that the effects were long-term and had a larger impact on women as well as the young.[73]

Schep sees social cooling has having three primary impacts on society.[74] Firstly, more self-censorship leads to people not saying things, clicking on things, or searching for things, despite neither legislation or nor policy preventing it. Secondly, it will hold back society but limiting people's perception of freedom to protest, stand up and rebel. Finally, it will lead to increased risk-avoidance.

A major challenge brought on by the concept of social cooling is that it is very difficult to measure the impact of the gradual creep towards all pervasive surveillance. The Snowden revelations were reported so widely across the world and offered such concrete proof of data collection by the state that it offered researchers the possibility to compare two different worlds: the pre- and post-Snowden world. We may only realise the impact of social cooling by looking back to the past, by which time it may be too late to do anything about it.

CONCLUSION

Data is now a valuable commodity. Although legislative bodies are finally waking up to the brave new world of data gathering, the ability of huge businesses to dodge regulations means that measures like GDPR are unlikely to have bite. Currently, individuals are made responsible for managing their own personal data, rather than restricting the power of data gathering entities. As a result, we risk facing a world where only the wealthy are able to afford to maintain their privacy. When compounded with the fact that data will often be stored forever we risk a profound where we are forever held accountable for past actions and ultimately may shift the online space from somewhere we can express our true selves and find community and belonging to a mechanism of control and conformity. The prevalence of data collection is concentrating immense power in the hands of large companies and government agencies, with most people excluded from the potential empowerment of the data revolution.

ENDNOTES

[1] Weigand A (2016) 'Data for the People', Basic Books
[2] https://www.ft.com/content/dd5e5514-198d-11e4-8730-00144feabdc0
[3] http://cyberlaw.stanford.edu/blog/2014/08/everyone-knows-privacy-about-power-now-what
[4] Angwin J (2014) Dragnet Nation, Times Books
[5] Solove, Daniel J., Privacy and Power: Computer Databases and Metaphors for Information Privacy. Stanford Law Review, Vol. 53, p. 1393, July 2001. Available at SSRN: https://ssrn.com/abstract=248300 p.1398
[6] Taddicken M, 'The 'Privacy Paradox' in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure', *Journal of Computer-Mediated Communication*, Volume 19, Issue 2, 1 January 2014, Pages 248–273
[7] Blank, Bolsover & Dubois. "A new privacy paradox: young people and privacy on social networks", April 2014, Global Cyber Security Capacity Centre
[8] Wolfie Christl (2017) "Corporate Surveillance in Everyday Life"
[9] Richards & King, "Three paradoxes of big data" - https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/66_StanLRevOnline_41_RichardsKing.pdf
[10] Richards & King, "Three paradoxes of big data" - https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/66_StanLRevOnline_41_RichardsKing.pdf p.45
[11] https://whoownsengland.org/
[12] Schwartz J, "Giving Web a memory cost its users privacy", 4th July 2001, New Your Times
[13] Wall street Journal (2010) "What they know", http://www.wsj.com/public/page/what-they-know-2010.html
[14] Christl & Spiekermann, "Networks of Control", 2016, p.46
[15] Christl & Spiekermann, "Networks of Control", 2016, p.46-75
[16] https://techcrunch.com/2017/06/05/apple-adds-a-tracker-blocker-to-desktop-safari/
[17] https://www.theverge.com/2017/9/14/16308138/apple-safari-11-advertiser-groups-cookie-tracking-letter
[18] Alex Hern, "No tracking, no revenue: Apple's privacy feature costs ad companies millions, The Guardian, 9th January 2017, https://www.theguardian.com/technology/2018/jan/09/apple-tracking-block-costs-advertising-companies-millions-dollars-criteo-web-browser-safari
[19] An excellent example is Better Blocker (https://better.fyi/)
[20] https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf
[21] https://www.ofcom.org.uk/about-ofcom/latest/media/media-releases/2015/time-spent-online-doubles-in-a-decade
[22] https://hackernoon.com/how-much-time-do-people-spend-on-their-mobile-phones-in-2017-e5f90a0b10a6
[23] Exploring News Apps and Location-Based Services on the Smartphone (2013) Amy Schmitz Weiss
[24] http://www.wired.co.uk/article/google-history-search-tracking-data-how-to-delete
[25] "Personal data" is defined broadly as "any information relating to an identified or identifiable natural person ('data subject')"
[26] There are no protections for the gathering of non-personal data
[27] http://blog.pritchettslaw.com/2016/05/obtaining-valid-consent-under-data.html
[28] https://www.theguardian.com/technology/2017/may/17/inquiry-launched-into-how-uk-parties-target-voters-through-social-media
[29] https://www.theguardian.com/technology/2018/feb/16/facebook-ordered-stop-collecting-user-data-fines-belgian-court
[30] Cate F, "The failure of fair information practice principles", in Winn J (ed.), *Consumer Protection in the Age of the Information Economy*, p.360
[31] Rubenstien I, "Big Data: the end of privacy or a new beginning", (2012) New York University Public Law and Legal Theory working papers, paper 357

[32] ICO, "GDPR Consent Guidance" - https://ico.org.uk/media/about-the-ico/consultations/2013551/draft-gdpr-consent-guidance-for-consultation-201703.pdf
[33] https://iconewsblog.org.uk/2017/08/16/consent-is-not-the-silver-bullet-for-gdpr-compliance/
[34] https://www.opensocietyfoundations.org/sites/default/files/data-brokers-in-an-open-society-20161121.pdf p.30
[35] GDPR, Recital 38 and 39

[36] https://www.opensocietyfoundations.org/sites/default/files/data-brokers-in-an-open-society-20161121.pdf p.23
[37] Rubenstien I, "Big Data: the end of privacy or a new beginning", (2012) New York University Public Law and Legal Theory working papers, paper 357
[38] Rubenstien I, "Big Data: the end of privacy or a new beginning", (2012) New York University Public Law and Legal Theory working papers, paper 357
[39] https://www.wired.com/story/yahoo-breach-three-billion-accounts/
[40] Duhigg C (2013) The Power of Habit
[41] https://medium.com/@powerb91/equifax-data-breach-due-to-negligence-170ed47d2a27
[42] https://www.digitaltrends.com/social-media/mark-zuckerberg-daughter-vr/
[43] https://www.theguardian.com/technology/2010/jan/11/facebook-privacy
[44] https://www.dailydot.com/layer8/online-privacy-data-ethics/
[45] Angwin J, "Dragnet Nation" – pages about medical/drug history site sharing
[46] Bergstrom A, 'Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses', Computers in Human Behavior 53 (2015) 419–426 p.419
[47] Baek, Y. M., Kim, E., & Bae, Y. (2014). My privacy is okay, but theirs is endangered: Why comparative optimism matters in online privacy concerns. Computers in Human Behavior, 31, 48–56.
[48] Barnes S, "A Privacy paradox: social networking in the US", 4th September 2006, First Monday, 11(9), http://firstmonday.org/article/view/1394/1312_2
[49] Blank, Bolsover & Dubois. "A new privacy paradox: young people and privacy on social networks", April 2014, Global Cyber Security Capacity Centre p.22
[50] Hoofnagle, King, Li & Truow, "How different are your adults from older adults when it comes to information privacy attitudes and policy", p.20
[51] Blank, Bolsover & Dubois. "A new privacy paradox: young people and privacy on social networks", April 2014, Global Cyber Security Capacity Centre p.25
[52] UK Parliament. 'The Big Data Dilemma' https://publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/46802.htm
[53] https://www.theguardian.com/commentisfree/2017/aug/07/the-guardian-view-on-data-protection-a-vital-check-on-power
[54] https://www.theguardian.com/commentisfree/2017/may/20/eu-right-to-take-on-facebook-fines-dont-protect-us-from-tech-giants
[55] Danah Boyd, "Networked Privacy", presented at the Personal Democracy Forum, 2011, New York
[56] Nissenbaum H & Barocas S (2014) "Big Data's End Run around anonymity and consent", in "Privacy, Big Data and the Public Good" (ed Lane, Stodden, Nissenbaum, Bender), Cambirdge University Press, p.61
[57] Mislove et al., "You are who you know: inferring user profiles in online social networks" p.255
[58] Hoofnagle, King, Li & Truow, "How different are your adults from older adults when it comes to information privacy attitudes and policy", p.20
[59] https://www.theguardian.com/media/2018/jan/09/toby-young-resigns-office-for-students
[60] https://www.wired.com/story/how-one-womans-digital-life-was-weaponized-against-her/
[61] EU Data Protection Directive 95/46/EC Art 12 & 28 - http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML
[62] C-131/12 - http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf
[63] https://venturebeat.com/2017/06/28/canadas-top-court-rules-google-can-be-forced-to-remove-search-results-worldwide/
[64] https://gizmodo.com/ominous-right-to-be-forgotten-case-with-global-conseque-1797067061
[65] https://www.theguardian.com/world/2017/feb/02/right-to-be-forgotten-online-suffers-setback-after-japan-court-ruling

[66] https://www.ft.com/content/c5d17a80-0910-11e4-8d27-00144feab7de#axzz3gL6XFlzC
[67] https://thenextweb.com/google/2016/09/12/google-removed-1-75-billion-websites-copyright-takedown-requests/
[68] https://transparencyreport.google.com/eu-privacy/overview
[69] https://www.socialcooling.com/
[70] http://www.abc.net.au/news/2017-09-07/social-cooling-are-you-self-conscious-about-what-you-click-on/8878948
[71] Guardian NSA files, https://www.theguardian.com/us-news/the-nsa-files
[72] Marthews & Tucker, "Government Surveillance and Internet Search Behaviour", (2017) Available at SSRN: https://ssrn.com/abstract=2412564
[73] Penney, Jon, Chilling Effects: Online Surveillance and Wikipedia Use (2016). Berkeley Technology Law Journal, Vol. 31, No. 1, p. 117, 2016. Available at SSRN: https://ssrn.com/abstract=2769645
[74] http://www.abc.net.au/news/2017-09-07/social-cooling-are-you-self-conscious-about-what-you-click-on/8878948