Privacy after Big Data

Compilation of Early Research

AMBER SINHA
ELONNAI HICKOK
ROHAN GEORGE
SCOTT MASON
SEAN MARTIN MCDONALD
VANYA RAKESH
VIPUL KHARBANDA

12th November, 2016

Research in the compilation has been supported by **Privacy International** with funding from **SIDA**, **IDRC**, **MacArthur Foundation**, and **Open Society Foundations**.

The article by **Sean Martin McDonald** is an extract from his study titled '**Ebola: A Big Data Disaster**', undertaken with support from the **Open Society Foundation, Ford Foundation,** and **Media Democracy Fund**, and published by the **CIS**.

Compiled by **Vanya Rakesh, Amber Sinha, and Sumandro Chattapadhyay**
Designed by **Saumyaa Naidu**

Privacy after Big Data

Evolving data science, technologies, techniques, and practices, including big data, are enabling shifts in how the public and private sectors carry out their functions and responsibilities, deliver services, and facilitate innovative production and service models to emerge. For example, in the public sector, the Indian government has considered replacing the traditional poverty line with targeted subsidies based on individual household income and assets. The my.gov.in platform is aimed to enable participation of the connected citizens, to pull in online public opinion in a structured manner on key governance topics in the country. The 100 Smart Cities Mission looks forwards to leverage big data analytics and techniques to deliver services and govern citizens within city sub-systems. In the private sector, emerging financial technology companies are developing credit scoring models using big, small, social, and fragmented data so that people with no formal credit history can be offered loans. These models promote efficiency and reduction in cost through personalization and are powered by a wide variety of data sources including mobile data, social media data, web usage data, and passively collected data from usages of IoT or connected devices.

These data technologies and solutions are enabling business models that are based on the ideals of 'less': cash-less, presence-less, and paper-less. This push towards an economy premised upon a foundational digital ID in a prevailing condition of absent legal frameworks leads to substantive loss of anonymity and privacy of individual citizens and consumers vis-a-vis both the state and the private sector. Indeed, the present use of these techniques run contrary to the notion of the 'sunlight effect' - making the individual fully transparent (often without their knowledge) to the state and private sector, while the algorithms and means of reaching a decision are opaque and inaccessible to the individual.

These techniques, characterized by the volume of data processed, the variety of sources data is processed from, and the ability to both contextualize - learning new insights from disconnected data points - and de-contextualize - finding correlation rather than causation - have also increased the value of all forms of data. In some ways, big data has made data exist on an equal playing field as far as monetisation and joining up are concerned. Meta data can be just as valuable to an entity as content data. As data science techniques evolve to find new ways of collecting, processing, and analyzing data - the benefits of the same are clear and tangible, while the harms are less clear, but significantly present.

Is it possible for an algorithm to discriminate? Will incorrect decisions be made based on data collected? Will populations be excluded from necessary services if they do not engage with certain models or do emerging models overlook certain populations? Can such tools be used to surveil individuals at a level of granularity that was formerly not possible and before a crime occurs? Can such tools be used to violate rights – for example target certain types of speech or groups online? And importantly, when these practices are opaque to the individual, how can one seek appropriate and effective remedy.

Traditionally, data protection standards have defined and established protections for certain categories of data. Yet, data science techniques have evolved beyond data protection principles. It is now infinitely harder to obtain informed consent from an individual when data that is collected can be used for multiple purposes by multiple bodies. Providing notice for every use is also more difficult – as is fulfilling requirements of data minimization. Some say privacy is dead in the era of big data. Others say privacy needs to be re-conceptualized, while others say protecting privacy now, more than ever, requires a 'regulatory sandbox' that brings together technical design, markets, legislative reforms, self regulation, and innovative regulatory frameworks. It also demands an expanding of the narrative around privacy – one that has largely been focused on harms such as misuse of data or unauthorized collection – to include discrimination, marginalization, and competition harms.

In this compilation we have put together a series of articles that we have developed as we explore the impacts – positive and negative – of big data. This includes looking at India's data protection regime in the context of big data, reviewing literature on the benefits of harms of big data, studying emerging predictive policing techniques that rely on big data, and analyzing closely the impact of big data on specific privacy principles such as consent. This is a growing body of research that we are exploring and is relevant to multiple areas of our work including privacy and surveillance. Feedback and comments on the compilation are welcome and appreciated.

ELONNAI HICKOK

Contents

Legal Landscape for Privacy

ELONNAI HICKOK AND VIPUL KHARBANDA

India is a signatory to the Universal Declaration on Human Rights (Article 12) and the International Convention on Civil and Political Rights (Article 17) – both of which recognize privacy as a fundamental right. Though a member and signatory of these conventions, India does not have laws which guarantee a right to privacy to its citizens. In order to fill this lacuna in the law, the Courts in India have tried to enforce a right to privacy in favour of its citizens through two main routes, viz. a recognition of a constitutional right to privacy which has been read as part of the rights to life and personal liberty as well as the freedom of expression and movement guaranteed under the Constitution; and a common law right to privacy which is available under tort law and has been borrowed primarily from American jurisprudence. It must be mentioned at the outset that the privacy is not a very strongly enforced right in India and there are a number of exceptions to the right to privacy which have been carved out by the Courts over a period of time, which we shall discuss later in this section.

The right to privacy was recognised as a constitutional principle in India for the first time by the Supreme Court in 1962 in the case of Kharak Singh v. Union of India [1], in the background of the right of the police to physically keep checks on people who are repeat offenders (also known as history-sheeters). Although a majority of three Judges in the case categorically denied the existence of a constitutional right to privacy as part of the right to life and personal liberty [2], two Judges disagreed with this analysis and held that

"...the right to personal liberty takes in not only a right to be free from restrictions placed on his movements, but also free from encroachments on his private life. It is true our Constitution does not expressly declare a right to privacy as a fundamental right, but the said right is an essential ingredient of personal liberty."

Later in the case of Govind v. State of M.P. [3], the Supreme Court in a similar factual background endorsed the view taken by the minority in Kharak Singh [4] and the right to privacy in India has become understood as an entrenched constitutional right under Indian law. In this case the Supreme Court discussed the right to privacy at length and debated the scope of and exceptions to this right as well. After discussing the scope of the right in the Indian context, the Court came to the conclusion that the right to privacy is not an absolute right and laid down three different tests which can be used to determine whether the right to privacy would be upheld/enforced in a given situation or not. These three tests are: (i) important countervailing interest which is superior, (ii) compelling state interest test, and (iii) compelling public interest. However, the Court later used phrases such as "reasonable restriction in public interest" and "reasonable restriction upon it for compelling interest of State" interchangeably which seems to suggest that the terms "compelling public interest" and "compelling state interest" used by the Court were being used synonymously and the Court does not draw any distinction between them. It is also important to note that the wider phrase "countervailing interest is shown to be superior" seems to suggest that it is possible, at least in theory, to have other interests apart from public interest or state interest also which could trump the right to privacy.

After the case of Govind v. State of M.P. [5], the right to privacy was firmly established as a fundamental right guaranteed to the citizens of India, but with a limited scope and a number of exceptions. We could discuss all the judgements that have sculpted the constitutional jurisprudence on privacy in Indian law, but for the sake of brevity and to avoid repetition we will just summarize the ratio from those cases in a few bullet points:

a. Reasonable restrictions can be imposed on the right to privacy in the interests of the sovereignty and integrity of India, the security of the State, friendly relations with foreign

States, public order, decency or morality, or in relation to contempt of court, defamation or incitement to an offence [6];

b. Reasonable restrictions can be imposed upon the right to privacy either in the interests of the general public or for the protection of the interests of any Scheduled Tribe [7];

c. The right to privacy can be restricted by procedure established by law which procedure would have to satisfy the test laid down in the Maneka Gandhi case [8].

d. The right can be restricted if there is an important countervailing interest which is superior [9];

e. It can be restricted if there is a compelling state interest to be served by doing so [10];

f. It can be restricted in case there is a compelling public interest to be served by doing so [11];

g. The Rajagopal tests - This case lays down three exceptions to the rule that a person's private information cannot be published, viz. i) person voluntarily thrusts himself into controversy or voluntarily raises or invites a controversy, ii) if publication is based on public records other than for sexual assault, kidnap and abduction, iii) there is no right to privacy for public officials with respect to their acts and conduct relevant to the discharge of their official duties. It must be noted that although the Court talks about public records, it does not use the term 'public domain' and thus it is possible that even if a document has been leaked in the public domain and is freely available, if it is not a matter of public record, the right to privacy can still be claimed in regard to it [12].

The scope and exceptions to the right to privacy both have developed over the years in various contexts through various cases which have laid down the scope of the right to privacy in specific fields, which we shall discuss below so as to give a brief overview of the privacy landscape in Indian laws.

Freedom of the Press

It is pretty obvious to anyone at first glance that the right to privacy would share a complicated relationship with the freedom of the press, which is protected by the fundamental right of freedom of expression. In India there is an institution called the Press Council of India which is a mechanism for the Press to regulate itself. The Press Council of India issues various Norms and Guidelines which address some of the thorny issues of journalistic conduct including the right to privacy of the subjects of their reporting. It must be noted that some of the principles enunciated in these guidelines have been adopted from those propounded by the Courts while discussing the issue of privacy and freedom of the Press. Although the guidelines do specifically regulate some of the areas of journalistic conduct regarding the privacy of the subjects, however these guidelines are not legally enforceable and therefore lack any a strong enough mechanism to ensure their compliance.

The guidelines state that "The Press shall not intrude or invade the privacy of an individual, unless outweighed by genuine overriding public interest, not being a prurient or morbid curiosity. So, however, that once a matter becomes a matter of public record, the right to privacy no longer subsists and it becomes a legitimate subject for comment by the Press and the media, among others. Special caution is essential in reports likely to stigmatise women" [13]. The guidelines make a distinction between the degree of privacy to be enjoyed by different people and under different circumstances, i.e. a representative or emissary of the public is not afforded the same degree of privacy as a private person. The guidelines also caution against identification of victims of sexual abuse, children of forcible marriages or illicit sexual unions, etc. They also ask journalists not to intrude into moments of personal grief of individuals [14].

Other than these guidelines by the Press Council of India, the only legislation of some impact which affects the issue of privacy and the media is the Juvenile Justice (Care and Protection)

of Children Act, 2000 which makes it an offence for the media to disclose the names, addresses or schools, or pictures of juveniles who are involved in a legal proceeding under the Act which would lead to their identification [15].

The issue of freedom of the press and privacy was discussed at length in the case of R. Rajagopal v. Union of India [16], (also known as the Auto Shanker case) where certain public officials approached the Supreme Court to stop the publication of an autobiography of a convict which contained embarrassing allegations regarding their dealings with the convict. It was argued that allowing the publication of these details/allegations would amount to a breach the right to privacy of these public officials. After a detailed discussion of the scope of the right to privacy, the Court agreed with the principle that a person's private information cannot be published but laid down three exceptions to this rule, viz. i) person voluntarily thrusts himself into controversy or voluntarily raises or invites a controversy, ii) if publication is based on public records other than for sexual assault, kidnap and abduction, iii) there is no right to privacy for public officials with respect to their acts and conduct relevant to the discharge of their official duties. It must be noted that although the Court talks about public records, it does not use the term 'public domain' and thus it is possible that even if a document has been leaked in the public domain and is freely available, if it is not a matter of public record, the right to privacy can still be claimed in regard to it.

The above stated principles were further clarified in the case of Indu Jain v. Forbes Incorporated [17], where a case was filed by Indu Jain in the Delhi High Court to stop Forbes magazine from featuring her family in the Forbes List of Indian Billionaires. After a discussion of the various authorities and cases on the issue the Court further clarified the principles relating to privacy and freedom of the press which are summarized below:

- In evaluating a relief to be granted in respect of a complaint against infraction of the right to privacy, the court has to balance the rights of the persons complaining of infraction of right to privacy against freedom of press and the right of public to disclosure of newsworthy information. Such consideration may entail the interest of the community and the court has to balance the proportionality of interfering with one right against the proportionality of impact by infraction of the other.
- Since public figures like public officials play an influential role in ordering society the citizen has a legitimate and substantial interest in the conduct of such persons and the freedom of press extends to engaging in uninhibited debate about the involvement of public figures in public issues and events.
- A public person or personage is one who by his standing, accomplishment, fame, mode of life or by adopting a profession or calling which gives the public a legitimate interest in his doings, affairs and character has so become a public figure and thereby relinquishes at least a part of his privacy.
- Public or general interest in the matter published has to be more than mere idle curiosity for it to be an exception to the right to privacy.
- The standard to be adopted for assessing as to whether the published material infracts the right to privacy of any individual is that of an ordinary man of common sense and prudence and not an out of ordinary or hyper-sensitive man.
- The publication has to be judged as a whole and news items, advertisements and published matter cannot be read without the accompanying message that is purported to be conveyed to public. Pre-publication censorship may not be countenanced in the scheme of the constitutional framework unless it is established that the publication has been made with reckless disregard for truth, publication shall not be normally prohibited.

Medical Privacy

Confidentiality and privacy are essential to all trusting relationships, such as that between patients and doctors. Moreover, in a healthcare context, patient confidentiality and the protection of privacy is the foundation of the doctor-patient relationship. Patients must feel comfortable sharing private information about their bodily functions, physical and sexual activities, and medical history. In areas where there exists a traditionally strong jurisprudence of confidentiality, such as health, finance, etc. Indian law follows the English common law principles of confidentiality which generally requires three elements to succeed, apart from contract, (i) the information itself must have the necessary quality of confidence about it, (ii) that information must have been imparted in circumstances importing an obligation of confidence, and (iii) there must be an unauthorized use of that information to the detriment of the party communicating it.

The medical profession in India is regulated and overseen by the Medical Council of India (MCI) which issues certain regulations from time to time to regulate the conduct of the medical professionals in India. The MCI's Code of Ethics Regulations, 2002 impose a duty on physicians to protect the confidentiality of patients including their personal and domestic lives, unless the law requires their revelation, or if there is a serious and identified risk to a specific person and / or community or notifiable disease. Similarly the Ethical Guidelines for Biomedical Research on Human Subjects released by the Indian Council for Medical Research in 2006 contain specific provisions relating to privacy of the research subjects [18].

Apart from the above regulations there have been a few important cases in the field of medical privacy which have further tried to clarify the scope of the right to privacy in the medical and health sector. The first really important case in this field was Mr. X v. Hospital Z, Supreme Court of India [19], where the Supreme Court had to strike a balance between a patients' right to privacy vis-a-vis the health and well being of third parties who may be at risk by coming into contact with the patient. The facts of this case were very interesting, a person who was engaged to be married was found to be HIV positive. This information was released by the doctor to the petitioner's family and through them to the family of his fiancée without the consent of the patient. The Supreme Court in this case held that a person could not invoke his "right to privacy" to prevent a doctor from disclosing his HIV-positive status to others. It was ruled that in respect of HIV-positive persons, the duty of confidentiality between the doctor and patient could be compromised in order to protect the health of other individuals. The Court held that in such a case disclosure by the Doctor could not be violative of either the rule of confidentiality or the patient's right of privacy as the person with whom the patient was likely to be married was saved in time by such disclosure. The Court held that:

"26. Right of Privacy may, apart from contract, also arise out of a particular specific relationship which may be commercial, matrimonial, or even political. As already discussed above, Doctor-patient relationship, though basically commercial, is, professionally, a matter of confidence and, therefore. Doctors are morally and ethically bound to maintain confidentiality. In such a situation, public disclosure of even true private facts may amount to an invasion of the Right of Privacy which may sometimes lead to the clash of person's "right to be let alone" with another person's right to be informed.

28. Having regard to the fact that the appellant was found to be HIV(+), its disclosure would not be violative of either the rule of confidentiality or the appellant's Right of Privacy as Ms. Akali with whom the appellant was likely to be married was saved in time by such disclosure, or else, she too would have been infected with the dreadful disease if marriage had taken place and consummated."

Expanding the logic of this case further, it could be argued that this principle of disclosure to the person at risk could be applicable to other communicable and life threatening diseases as well, while a conservative opinion would be that this principle should only be applied to HIV+ cases. It may also be noted that the Court in this case did not discuss whether

disclosure of such a disease may be made to only the persons at risk or to the public in general, although it would be safe to say that since most cases as well as medical ethics require the doctors to keep the patient's health records private, it is highly unlikely that 'disclosure' can be made to the public in general.

The other landmark decision in the realm of medical privacy is the case of Sharda v. Dharmpal, where the basic question was whether a party to a divorce proceeding can be compelled to a medical examination. The wife in the divorce proceeding refused to submit herself to medical examination to determine whether she was of unsound mind on the ground that such an act would violate her right to personal liberty. Discussing the balance between protecting the right to privacy and other principles that may be involved in matrimonial cases such as the 'best interest of the child' in case child custody is also in issue, the Court held:

"75. If the nature of the information relates directly to the well-being of the child or to the parent's ability to adequately care for child, and the court believes the child is potentially in danger, courts are likely to admit the information despite a patient's expectation of confidentiality. There are two competing interests involved when a court determines whether to compel discovery of a patient-litigant's mental health records over his objection in a child custody dispute. The first involves the privacy, confidentiality and privilege expectation of both the patient and the treating mental health professional in those communications. The second involves the application of the best interests of the child(ren) standard. Virtually every jurisdiction in the United States makes a child custody determination based upon the "best interest of the child."

It is interesting to note that in the previous cases it was a balance between the competing rights of two people, whereas in this case it was more about the right of one person vis-a-vis the best interest of a child, which incidentally has almost always been the deciding factor in Indian cases involving child custody.

Financial Privacy

Financial privacy involves the protection of consumers from unlawful access to financial accounts by private and public bodies, and the unlawful disclosure, sharing, or commercial use of financial information. Just as in the case of medical privacy, Indian law does cast a duty on bankers to protect the privacy of their customers. This duty of confidentiality is an extra layer of protection when it comes to financial information of individuals in addition to their right to privacy. Both these principles were used by the banks in one of the most important cases in the field of financial privacy, i.e. District Registrar and Collector, Hyderabad v. Canara Bank and others [21], where a provision of law which allowed the person inspecting the documents to also seize and impound the documents was challenged by the banks. The provision also extended this power of inspection to include not only public officers but also to citizens and banks. It was challenged inter alia, on the ground that it intruded into the privacy and property of individuals. Considering the issue of allowing such inspections at banks which held the private documents of their customers or copies of such private documents, the question before the Court was whether disclosure of the contents of the documents by the banks would amount to a breach of confidentiality and would, therefore, be violative of privacy rights of their customers? Discussing this issue the Court held as follows:

"It cannot be denied that there is an element of confidentiality between a Bank and its customers in relation to the latter's banking transactions...Once we have accepted in Govind and in latter cases that the right to privacy deals with 'persons and not places', the documents or copies of documents of the customer which are in Bank, must continue to remain confidential vis-a-vis the person, even if they are no longer at the customer's house and have been voluntarily sent to a Bank. If that be the correct view of the law, we cannot accept the line of Miller in which the Court proceeded on the basis that the right to privacy is referable to the right of 'property' theory. Once that is so, then unless there is some probable

or reasonable cause or reasonable basis or material before the Collector for reaching an opinion that the documents in the possession of the Bank tend, to secure any duty or to prove or to lead to the discovery of any fraud or omission in relation to any duty, the search or taking notes or extracts therefore, cannot be valid. The above safeguards must necessarily be read into the provision relating to search and inspection and seizure so as to save it from any unconstitutionality."

Secondly, the provision was also struck down because it enabled the Collector to authorize 'any person' whatsoever to inspect, to take notes or extracts from the papers in the public office since this was considered by the Court to be excessive delegation as there were no guidelines regulating this and the it allowed the facts relating to the customer's privacy to reach non-governmental persons and would, on that basis, be an unreasonable encroachment into the customer's rights. Therefore, the Court held this provision to be unconstitutional and struck it down.

Although banks are required to maintain confidentiality and thereby protect the privacy of their customers in their ordinary course of business, however sometime the duty to protect the right to privacy of their customers can come in direct conflict with the discharge of their functions. This situation has arisen a number of times when banks have sought to publish the photographs and information of wilful defaulters in newspapers, this practice has sometimes been objected to by the defaulters whose information is sought to be published on the ground that such publication would violate their right to privacy. There is currently a difference of opinion between various courts of law (i.e. High Courts in different States (provinces) have given different and opposing opinions on this issue). Some Courts have held that Banks are allowed to publish photographs of the defaulters even though it may violate the right to privacy of the defaulters since it would serve the interest of the bank and the economy as a whole by ensuring better recovery of bad loans [22]. On the other hand it has also been held by other Courts that (government owned) banks have the power to realize their dues only in a manner authorized by law and there is no provision in law which allows the bank to publish the photographs of defaulters in newspapers. It further held that such an action by the banks would violate the right to privacy of the individuals [23].

Apart from the general duty of bankers to maintain confidentiality, there are certain legislations which also touch upon the need to maintain secrecy in financial transactions. The Credit Information Companies (Regulation) Act, 2005 and the Regulations thereunder which provide that specific instances under which credit information of individuals may be released by the credit information companies. Further, as far as financial institutions owned by the government are concerned the Public Financial Institutions (Obligation as to Fidelity and Secrecy) Act, 1983 [24] as well as the State Bank of India Act, 1955 [25] provide that these institutions are prohibited from divulging any information relating to the affairs of its clients except in accordance with laws of practice and usage. To enforce this all their employees must take an oath of secrecy before carrying out their duties.

ENDNOTES

[1] http://www.judis.nic.in/supremecourt/imgs1.aspx?filename=3641.

[2] Article 21 of the Constitution of India, 1950. However in the later case of PUCL v. Union of India (discussed later), the Supreme Court insisted that all seven Judges had interpreted Article 21 to include the right to privacy.

[3] http://www.judis.nic.in/supremecourt/imgs1.aspx?filename=6014.

[4] Supra Note 1.

[5] Supra Note 3.

[6] Article 19(2) of the Constitution of India, 1950.

[7] Article 19(5) of the Constitution of India, 1950.

[8] Maneka Gandhi v. Union of India, Supreme Court of India, WP No. 231 of 1977, dated 25-01-1978. The test laid down in this case is universally considered to be that the procedure established by law which restricts the fundamental right should be just, fair and reasonable.

[9] Govind v. State of M.P., Supreme Court of India, WP No. 72 of 1970, dated 18-03-1975.

[10] Id.

[11] Id.

[12] R. Rajagopal v. Union of India, Supreme Court of India, dated 7-10-1994. These tests have been listed as one group since they are all applicable in the specific context of publication of private information.

[13] Principles and Ethics for Journalists, Press Council of India, http://presscouncil.nic.in/Content/62_1_PrinciplesEthics.aspx.

[14] Id.

[15] Section 21, Juvenile Justice (Care and Protection of Children) Act, 2000.

[16] http://www.judis.nic.in/supremecourt/imgs1.aspx?filename=11212

[17] http://lobis.nic.in/dhc/GM/judgement/25-01-2010/GM12102007S21722006.pdf

[18] Ethical Guidelines for Biomedical Research on Human Subjects. (2006) Indian Council of Medical Research New Delhi.

[19] http://www.judis.nic.in/supremecourt/imgst.aspx?filename=19994.

[20] http://www.judis.nic.in/supremecourt/imgs1.aspx?filename=19039

[21] http://www.judis.nic.in/supremecourt/imgs1.aspx?filename=26571

[22] K.J. Doraisamy v. Asst. General Manager, http://judis.nic.in/judis_chennai/qrydisp.aspx?filename=8673.

[23] Venu P.R. v. Assistant General Manager, SBI and another, 2013 (132) AIC 612 (Ker.H.C.).

[24] Sections 3 and 4, Public Financial Institutions (Obligation as to Fidelity and Secrecy) Act, 1983.

[25] Section 45, State Bank of India Act, 1955.

Nature of Knowledge

SCOTT MASON

Introduction

In 2008 Chris Anderson infamously proclaimed the 'end of theory'. Writing for *Wired* Magazine, Anderson predicted that the coming age of Big Data would create a 'deluge of data' so large that the scientific methods of hypothesis, sampling and testing would be rendered 'obsolete' [1]. For him and others, the hidden patterns and correlations revealed through Big Data analytics enable us to produce objective and actionable knowledge about complex phenomena not previously possible using traditional methodologies. As Anderson himself put it, '*there is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot*' [2] .

In spite of harsh criticism of Anderson's article from across the academy, his uniquely (dis) utopian vision of the scientific utility of Big Data has since become increasingly mainstream with regular interventions from politicians and business leaders evangelising about Big Data's potentially revolutionary applications. Nowhere is this bout of data-philia more apparent than in India where the governments recently announced the launch of 'Digital India', a multi-million dollar project which aims to harness the power of public data to increase the efficiency and accessibility of public services [3]. In spite of the ambitious promises associated with Big Data however, many theorists remain sceptical about its practical benefits and express concern about its potential implications for conventional scientific epistemologies. For them the increased prominence of Big Data analytics in science does not signal a paradigmatic transition to a more enlightened data-driven age, but a hollowing out of the scientific method and an abandonment of casual knowledge in favour of shallow correlative analysis. In response, they emphasise the continued importance of theory and specialist knowledge to science, and warn against what they see as the uncritical adoption of Big Data in public policy-making [4]. In this article I will examine the challenges posed by Big Data technologies to established scientific epistemologies as well as the possible implications of these changes for public-policymaking. Beginning with an exploration of some of the ways in which Big Data is changing our understanding of scientific research and knowledge, I will argue that claims that Big Data represents a new paradigm of scientific inquiry are predicated upon a number of implicit assumptions about the nature of knowledge. Through a critic of these assumptions I will highlights some of the potential risks that an over-reliance on Big Data analytics poses for public policy-making, before finally making the case for a more nuanced approach to Big Data, which emphasises the continued importance of theory to scientific research.

Big Data: The Fourth Paradigm?

"*Revolutions in science have often been preceded by revolutions in measurement*".

In his book the *Structure of Scientific Revolutions* Kuhn describes scientific paradigms as 'universally recognized scientific achievements that, for a time, provide model problems and solutions for a community of researchers' [5]. Paradigms as such designate a field of intelligibility within a given discipline, defining what kinds of empirical phenomena are to be observed and scrutinized, the types of questions which can be asked of those phenomena, how those questions are to be structured as well as the theoretical frameworks within which the results can be analysed and interpreted. In short, they 'constitute an accepted way of interrogating the world and synthesizing knowledge common to a substantial proportion of researchers in a discipline at any one moment in time' [6]. Periodically however, Kuhn

argues, that these paradigms can become destabilised by the development of new theories or the discovery of anomalies that cannot be explained through reference to the dominate paradigm. In such instances Kuhn claims, the scientific discipline is thrown into a period of 'crisis', during which new ideas and theories are proposed and tested, until a new paradigm is established and gains acceptance from the community.

More recently computer scientists Jim Gray, adopted and developed Kuhn's concept of the paradigm shift, charting history of science through the evolution of four broad paradigms, experimental science, theoretical science, computational science and exploratory science [7]. Unlike Kuhn however, who proposed that paradigm shifts occur as the result of anomalous empirical observations which scientists are unable to account for within the existing paradigm, Gray suggested that transitions in scientific practice are in fact primarily driven by advances and innovations in methods of data collection and analysis. The emergence of the *experimental paradigm* according to Gray can therefore be traced back to the ancient Greece and China when philosophers began to describe their empirical observations using natural rather an spiritual explanations. Likewise, the transition to the *theoretical paradigm* of science can be located in the 17th Century during which time scientists began to build theories and models which made generalizations based upon their empirical observations. Thirdly, Gray identifies the emergence of a *computational paradigm* in the latter part of the 20th Century in which advanced techniques of simulation and computational modelling were developed to help solve equations and explore fields of inquiry such as climate modelling which would have been impossible using experimental or theoretical methods. Finally, Gray proposed that we are today witnessing a transition to a 'fourth paradigm of science', which he termed the *exploratory paradigm*. Although also utilising advanced computational methods, unlike the previous computational paradigm which developed programs based upon established rules and theories, Gray suggested that within this new paradigm, scientists begin with the data itself; designing programs to mine enormous databases in the search for correlations and patterns; in effect using the data to discover the rules [8].

The implications of this shift are potentially significant for the nature of knowledge production, and are already beginning to be seen across a wide range of sectors. In the retail sector for example, data mining and algorithmic analysis are already being used to help predict items that a customers may wish to purchase based upon previous shopping habits [9]. Here, unlike with traditional research methodologies the analysis does not presuppose or hypothesise a relationship between items which it then attempts to prove through a process of experimentation, instead the relationships are identified inductively through the processing and reprocessing of vast quantities of data alone. By starting with the data itself, Big Data analysts circumvent the need for predictions or hypothesis about what one is likely to find, as Dyche observes '*mining Big Data reveals relationships and patterns that we didn't even know to look for*' [10]. Similarly, by focussing primarily on the search for correlations and patterns as opposed to causation Big Data analysts also reject the need interpretive theory to frame the results instead researchers claim the outcomes are inherently meaningful and interpretable by anyone without the need for domain specific or contextual knowledge. For example, Joh observes how Big Data is being used in policing and law enforcement to help make better decisions about the allocation of police resources. By looking for patterns in the crime data they are able to make accurate predictions about the localities and times in which crimes are most likely to occur and dispatch their officers accordingly [11]. Such analysis according to Big Data proponents requires no knowledge of the cause of the crime, nor the social or cultural context within which it is being perpetrated, instead predictions and assessments are made purely on the basis of patterns and correlations identified within the historical data by statistical modelling.

In summary then, Gray's exploratory paradigm represents a radical inversion of the deductive scientific method, allowing researchers to derive insights directly from the data itself without the use of hypothesis or theory. Thus it is claimed, by enabling the collection and analysis of datasets of unprecedented scale and variety Big Data allows analysts to 'let the data speak

for itself' [12], providing exhaustive coverage of social phenomena, and revealing correlations that are inherently meaningful and interpretable by anyone without the need for specialised subject knowledge or theoretical frameworks.

For Gray and others this new paradigm is made possible only by the recent exponential increase in the generation and collection of data as well as the emergence of new forms of data science, known collectively as "Big Data". For them the 'deluge of data' produced by the increase in the number of internet enabled devices as well as the nascent development of the internet of things, presents scientists and researchers with unprecedented opportunities to utilise data in new and innovative way to develop new insights across a wide range of sectors, many of which would have been unimaginable even 10 years ago. Furthermore, advances in computational and statistical methods as well as innovations in data visualization and methods of linking datasets, mean that scientist can now utilise the data available to its full potential or as professor Gary King quipped '*Big Data is nothing compared to a big algorithm*' [13].

These developments in statistical and computational analysis combined with the velocity variety and quantity of data available to analysts have therefore allowed scientists to pursue new types of research, generating new forms of knowledge and facilitating a radical shift in how we think about "science" itself. As Boyd and Crawford note, '*Big Data [creates] a profound change at the levels of epistemology and ethics. Big Data reframes key questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and the categorization of reality . . . [and] stakes out new terrains of objects, methods of knowing, and definitions of social life*' [14]. For many these changes in the nature of knowledge production provide opportunities to improve decision-making, increase efficiency, encourage innovation across a broad range of sectors from healthcare and policing to transport to international development [15]. For others however, many of the claims of Big Data are premised upon some questionable methodological and epistemological assumptions, some of which threat to impoverish the scientific method and undermine scientific rigour [16].

Assumptions of Big Data

Given its bold claims the allure of Big Data in both the public and privates sectors is perhaps understandable. However despite the radical and rapid changes to research practice and methodology, there has nevertheless seemingly been a lack of reflexive and critical reflection concerning the epistemological implications of the research practices used in Big Data analytics. And yet implicit within this vision of the future of scientific inquiry lie a number of important and arguably problematic epistemological and ontological assumptions, most notably;

- Big Data can provide comprehensive coverage of phenomenon, capturing all relevant information.

- Big Data does not require hypothesis, a priori theory, or models to direct the data collection or research questions.

- Big Data analytics do not require theoretical framing in order to be interpretable. The data is inherently meaningful transcending domain specific knowledge and can be understood be anyone.

- Correlative knowledge is sufficient to make accurate predictions and guide policy decisions.

For many, these assumptions are highly problematic and call into question the claims that Big Data makes about itself. I will now look at each one in turn before proposing there possible implications for Big Data in Policy-making.

Firstly, whilst Big Data may appear to be exhaustive in its scope, it can only be considered to be so in the context of the particular ontological and methodological framework chosen by the researcher. No data set however large can scrutinize all information relevant to a given

phenomenon. Indeed, even if it were somehow possible to capture all relevant quantifiable data within a specific domain, Big Data analytics would still be unable to fully account for the multifarious variables which are unquantifiable or undatafiable. As such Big Data does not provide an omnipresent 'gods-eye view', instead much like any other scientific sample it must be seen to provide the researcher with a singular and limited perspective from which he or she can observe a phenomenon and draw conclusions. It is important to recognise that this vantage point provides only one of many possible perspectives, and is shaped by the technologies and tools used to collect the data, as well as the ontological assumptions of the researchers. Furthermore, as with any other scientific sample, it is also subject to sampling bias and is dependent upon the researcher to make subjective judgements about which variables are relevant to the phenomena being studied and which can be safely ignored.

Secondly, claims by Big Data analysts to be able to generate insights directly from the data, signals a worrying divergence from deductive scientific methods which have been hegemonic within the natural sciences for centuries. For Big Data enthusiasts such as Prensky, '*scientists no longer have to make educated guesses, construct hypotheses and models, and test them with data-based experiments and examples. Instead, they can mine the complete set of data for patterns that reveal effects, producing scientific conclusions without further experimentation*' [17]. Whereas, deductive reasoning begins with general statements or hypotheses and then proceeds to observe relevant data equipped with certain assumptions about what should be observed if the theory is to be proven valid; inductive reasoning conversely begins with empirical observations of specific examples from which it attempts to draw general conclusions. The more data collected the greater the probability that the general conclusions generated will be accurate, however regardless of the quantity of observations no amount of data can ever conclusively prove causality between two variables, since it is always possible that my conclusions may in future be falsified by an anomalous observation. For example, a researcher who had only ever observed the existence of white swans may reasonably draw the conclusion that 'all swans are white', whilst they would be justified in making such a claim, it would nevertheless be comprehensively disproven the day a black swan was discovered. This is what David Hume called the 'problem of induction' [18] and strikes at the foundation of Big Data claims to be able to provide explanatory and predictive analysis of complex phenomena, since any projections made are reliant upon the 'principle of uniformity of nature', that is the assumption that a sequence of events will always occur as it has in the past. As a result, although Big Data may be well suited to providing detailed descriptive accounts of social phenomena, without theoretical grounding it nevertheless remains unable to prove casual links between variables and therefore is limited in its ability to provide robust explanatory conclusions or give accurate predictions about future events.

Finally, just as Big Data enthusiasts claim that theory or hypotheses are not needed to guide data collection, so too they insist human interpretation or framing is no longer required for the processing and analysis of the data. Within this new paradigm therefore, 'the data speaks for itself' [19], and specialised knowledge is not needed to interpret the results which are now supposedly rendered comprehensible to anyone with even a rudimentary grasp of statistics. Furthermore, the results we are told are inherently meaningful, transcending culture, history or social context and providing pure objective facts uninhibited by philosophical or ideological commitments.

Initially inherited from the natural sciences, this radical form of empiricism thus presupposes the existence of an objective social reality occupied by static and immutable entities whose properties are directly determinable through empirical investigation. In this way, Big Data reduces the role of social science to the perfunctory calculation and analysis of the mechanical processes of pre-formed subjects, in much the same way as one might calculate the movement of the planets or the interaction of balls on a billiard table. Whilst proponents of Big Data claim that such an approach allows them to produce objective knowledge, by cleansing the data of any kind of philosophical or ideological commitment, it nevertheless has the effect of restricting both the scope and character of social scientific inquiry; projecting onto the field of social research meta-theoretical commitments that have long

been implicit in the positivist method, whilst marginalising those projects which do not meet the required levels of scientificity or erudition.

This commitment to an empiricist epistemology and methodological monism is particularly problematic in the context of studies of human behaviour, where actions cannot be calculated and anticipated using quantifiable data alone. In such instances, a certain degree of qualitative analysis of social, historical and cultural variables may be required in order to make the data meaningful by embedding it within a broader body of knowledge. The abstract and intangible nature of these variables requires a great deal of expert knowledge and interpretive skill to comprehend. It is therefore vital that the knowledge of domain specific experts is properly utilized to help 'evaluate the inputs, guide the process, and evaluate the end products within the context of value and validity' [20].

Despite these criticisms however, Big Data is perhaps unsurprisingly increasingly becoming popular within the business community, lured by the promise of cheap and actionable scientific knowledge, capable of making their operations more efficient reducing overheads and producing better more competitive services. Perhaps most alarming from the perspective of Big Data's epistemological and methodological implications however, is the increasingly prominent role Big Data is playing in public policy-making. As I will now demonstrate, whilst Big Data can offer useful inputs into public policy-making processes, the methodological assumptions implicit within Big Data methodologies problems pose a number of risks to the effectiveness as well as the democratic legitimacy of public policy-making. Following an examination of these risks I will argue for a more reflexive and critical approach to Big Data in the public sector.

Big Data and Policy-Making: Opportunities and Risks

In recent year Big Data has begun to play an increasingly important role in public policy-making. Across the global, government funded projects designed to harvest and utilise vast quantities of public data are being developed to help improve the efficiency and performance of public services as well as better inform policy-making processes. At first glance, Big Data would appear to be the holy-grail for policy-makers - enabling truly evidence-based policy-making, based upon pure and objective facts, undistorted by political ideology or expedience. Furthermore, in an era of government debt and diminishing budgets, Big Data promises not only to produce more effective policy, but also to deliver on the seemingly impossible task of doing more with less, improving public services whilst simultaneously reducing expenditure.

In the Indian context, the government's recently announced 'Digital India' project promises to harness the power of public data to help modernise Indian's digital infrastructure and increase access to public services. The use of Big Data is seen as being central to the project's success, however, despite the commendable aspirations of Digital India, many commentators remain sceptical about the extent to which Big Data can truly deliver on its promises of better more efficient public services, whilst others have warned of the risk to public policy of an uncritical and hasty adoption of Big Data analytics [21]. Here I argue that the epistemological and methodological assumptions which are implicit within the discourse around Big Data threaten to undermine the goal of evidence based policy-making, and in the process widen already substantial digital divides.

It has long been recognised that science and politics are deeply entwined. For many social scientists the results of social research can be never entirely neutral, but are conditioned by the particular perspective of the researcher. As Shelia Jasanoff observed, 'Most thoughtful advisers have rejected the facile notion that giving scientific advice is simply a matter of ***speaking truth to power***. It is well recognized that in thorny areas of public policy, where certain knowledge is difficult to come by, science advisers can offer at best educated guesses and reasoned judgments, not unvarnished truth' [22]. Nevertheless, 'unvarnished truth' is precisely what Big Data enthusiasts claim to be able to provide. For them the capacity of Big Data to derive results and insights directly from the data without any need for human framing, allows policy-makers to incorporate scientific knowledge directly into their decision-

making processes without worrying about the 'philosophical baggage' usually associated with social scientific research.

However, in order to be meaningful, all data requires a certain level of interpretative framing. As such far from cleansing science of politics, Big Data simply acts to shift responsibility for the interpretation and contextualisation of results away from domain experts - who possess the requisite knowledge to make informed judgements regarding the significance of correlations - to bureaucrats and policy-makers, who are more susceptible to emphasise those results and correlations which support their own political agenda. Thus whilst the discourse around Big Data may promote the notion of evidence based policy-making, in reality the vast quantities of correlations generated by Big Data analytics act simply to broaden the range of 'evidence' from which politician can chose to support their arguments; giving new meaning to Mark Twain's witticism that there are 'lies, damn lies, and statistics'.

Similarly, for many an over-reliance on Big Data analytics for policy-making, risks leading to public policy which is blind to the unquantifiable and intangible. As already discussed above, Big Data's neglect of theory and contextual knowledge in favour of strict empiricism marginalises qualitative studies which emphasise the importance of traditional social scientific categories such as race, gender, and religion, in favour of a purely quantitative analysis of relational data. For many however consideration of issues such as gender, race, and religious sensitivity can be just as important to good public policy-making as quantitative data; helping to contextualise the insights revealed in the data and provide more explanatory accounts of social relations. They warn that neglect of such considerations as part of policy-making processes can have significant implications for the quality of the policies produced [23]. Firstly, although Big Data can provide unrivalled accounts of "what" people do, without a broader understanding of the social context in which they act, it fundamentally fails to deliver robust explanations of "why" people do it. This problem is especially acute in the case of public policy-making since without any indication of the motivations of individuals, policy-makers can have no basis upon which to intervene to incentivise more positive outcomes. Secondly, whilst Big Data analytics can help decision-makers to design more cost-effective policy, by for example ensuring better use of scarce resources; efficiency and cost-effectiveness are not the only metrics by which good policy can be judged. Public policy regardless of the sector must consider and balance a broad range of issues during the policy process including matters such as race, gender issues and community relations. Normative and qualitative considerations of this kind are not subject to a simplistic 1-0 quantification but instead require a great deal of contextual knowledge and insight to navigate successfully

Finally, to the extent that policy-makers are today attempting to harvest and utilise individual citizens personal data as direct inputs for the policy-making process, Big Data driven policy can in a very narrow sense be considered to offer a rudimentary form of direct democracy. At first glance this would appear to help to democratise political participation allowing public services to become automatically optimised to better meet the needs and preferences of citizens without the need for direct political participation. In societies such as India however, where there exist high levels of inequality in access to information and communication technologies, there remain large discrepancies in the quantities of data produced by individuals. In a Big Data world in which every byte of data is collected, analysed and interpreted in order to make important decisions about public services therefore, those who produce the greatest amounts of data, are better placed to have their voices heard the loudest, whilst those who lack access to the means to produce data risk becoming disenfranchised, as policy-making processes become configured to accommodate the needs and interests of a privilege minority. Similarly, using user generated data as the basis for policy decisions also leaves systems vulnerable to coercive manipulation. That is, once it has become apparent that a system has been automated on the basis of user inputs, groups or individuals may change their behaviour in order to achieve a certain outcome. Given these problems it is essential that in seeking to utilise new data resources for policy-making, we avoid an uncritical adoption of Big Data techniques, and instead as I argue below encourage a more balanced and nuanced approach to Big Data.

Data-Driven Science: A More Nuanced Approach?

Although an uncritical embrace of Big Data analytics is clearly problematic, it is not immediately obvious that a stubborn commitment to traditional knowledge-driven deductive methodologies would necessarily be preferable. Whilst deductive methods have formed the basis of scientific inquiry for centuries, the particular utility of this approach is largely derived from its ability to produce accurate and reliable results in situations where the quantities of data available are limited. In an era of ubiquitous data collection however, an unwillingness to embrace new methodologies and forms of analysis which maximise the potential value of the volumes of data available would seem unwise.

For Kitchen and others however, it is possible to reap the benefits of Big Data without comprising scientific rigour or the pursuit of casual explanations. Challenging the 'either or' propositions which favour either scientific modelling and hypothesis or data correlations, Kitchen instead proposes a hybrid approach which utilises the combined advantages of inductive, deductive and so-called 'abductive' reasoning, to develop theories and hypotheses directly from the data [24]. As Patrick W. Gross, commented 'In practice, the theory and the data reinforce each other. It's not a question of data correlations versus theory. The use of data for correlations allows one to test theories and refine them' [25].

Like the radical empiricism of Big Data, 'data-driven science' as Kitchen terms it, introduces an aspect of inductivism into the research design, seeking to develop hypotheses and insights 'born from the data' rather than 'born from theory'. Unlike the empiricist approach however, the identification of patterns and correlations is not considered the ultimate goal of the research process. Instead these correlations simply form the basis for new types of hypotheses generation, before more traditional deductive testing is used to assess the validity of the results. Put simply therefore, rather than interpreting data deluge as the 'end of theory', data-driven science instead attempts to harness its insights to develop new theories using alternative data-intensive methods of theory generation.

Furthermore unlike new empiricism, data is not collected indiscriminately from every available source in the hope that sheer size of the dataset will unveil some hidden pattern or insight. Instead, in keeping with more conventional scientific methods, various sampling techniques are utilised, 'underpinned by theoretical and practical knowledge and experience as to whether technologies and their configurations will capture or produce appropriate and useful research material' [26]. Similarly analysis of the data once collected does not take place within a theoretical vacuum, nor are all relationships deemed to be inherently meaningful; instead existing theoretical frameworks and domain specific knowledge are used to help contextualise and refine the results, identifying those patterns that can be dismissed as well as those that require closer attention.

Thus for many, data-driven science provides a more nuanced approach to Big Data allowing researchers to harness the power of new source of data, whilst also maintaining the pursuit of explanatory knowledge. In doing so, it can help to avoid the risks of uncritical adoption of Big Data analytics for policy-making providing new insights but also retaining the 'regulating force of philosophy'.

Conclusion

Since the publication of the *Structure of Scientific Revolutions*, Kuhn's notion of the paradigm has been widely criticised for producing a homogenous and overly smooth account of scientific progress, which ignores the clunky and often accidental nature of scientific discovery and innovation. Indeed the notion of the 'paradigm shift' is in many ways in typical of a self-indulgent and somewhat egotistical tendency amongst many historians and theorists to interpret events contemporaneous to themselves as in some way of great historical significance. Historians throughout the ages have always perceived themselves as living through periods of great upheaval and transition. In actual fact as has been noted by many, history and the history of science in particular rarely advances in a linear or predictable way, nor can progress when it does occur be so easily attributed to specific

technological innovations or theoretical developments. As such we should remain very sceptical of the claims that Big Data represents a historic and paradigmatic shift in scientific practice. Such claims exhibit more than a hint of technological determinism and often ignore the substantial limitations to Big Data analytics. In contrast to these claims, it is important to note that technological advances alone do not drive scientific revolutions; the impact of Big Data will ultimately depend on how we decide to use it as well as the types of questions we ask of it.

Big Data holds the potential to augment and support existing scientific practices, creating new insights and helping to better inform public policy-making processes. However, contrary to the hyperbole surrounding its development, Big Data does not represent a sliver-bullet for intractable social problems and if adopted uncritically and without consideration of its consequences, Big Data risks not only to diminishing scientific knowledge but also jeopardising our privacy and creating new digital divides. It is critical therefore that we see through the hyperbole and headlines to reflect critically on the epistemological consequences of Big Data as well as its implications for policy making, a task unfortunately which in spite of the pace of technological change is only just beginning.

BIBLIOGRAPHY

Anderson C (2008) The end of theory: The data deluge makes the scientific method obsolete. Wired, 23 June 2008. Available at: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory (accessed 31 October 2015).

Bollier D (2010) The Promise and Peril of Big Data. The Aspen Institute. Available at: http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf (accessed 19 October 2015).

Bowker, G., (2013) The Theory-Data Thing, International Journal of Communication 8 (2043), 1795-1799

Boyd D and Crawford K (2012) Critical questions for big data. Information, Communication and Society 15(5): 662-679.

Cukier K (2010) Data, data everywhere. The Economist, 25 February (accessed 5 November 2015).

Department of Electronics and Information Technology (2015) Digital India, [ONLINE] Available at: http://www.digitalindia.gov.in/. [Accessed 13 December 15].

Dyche J (2012) Big data 'Eurekas!' don't just happen, Harvard Business Review Blog. 20 November. Available at: http://blogs.hbr.org/cs/2012/11/eureka_doesnt_just_ happen.html

Hey, T., Tansley, S., and Tolle, K (eds)., (2009) The Fourth Paradigm: Data-Intensive Scientific Discovery, Redmond: Microsoft Research, pp. xvii-xxxi.

Hilbert, M. Big Data for Development: From Information- to Knowledge Societies (2013). Available at SSRN: http://ssrn.com/abstract=2205145

Hume, D., (1748), Philosophical Essays Concerning Human Understanding (1 ed.). London: A. Millar.

Jasanoff, S., (2013) Watching the Watchers: Lessons from the Science of Science Advice, Guardian 8 April 2013, available at: http://www.theguardian.com/science/political-science/2013/apr/08/lessons-science-advice

Joh. E, 'Policing by Numbers: Big Data and the Fourth Amendment', Washington Law Review, Vol. 85: 35, (2014) https://digital.law.washington.edu/dspacelaw/bitstream/handle/1773.1/1319/89WLR0035.pdf?sequence=1;

Kitchen, R (2014) Big Data, new epistemologies and paradigm shifts, Big Data & Society, April-June 2014: 1-12

Kuhn T (1962) The Structure of Scientific Revolutions. Chicago: University of Chicago Press.

Mayer-Schonberger V and Cukier K (2013) Big Data: A Revolution that Will Change How We Live, Work and Think. London: John Murray

McCue, C., Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis, Butterworth-Heinemann, (2014)

Morris, D. Big data could improve supply chain efficiency-if companies would let it, Fortune, August 5 2015, http://fortune.com/2015/08/05/big-data-supply-chain/

Prensky M (2009) H. sapiens digital: From digital immigrants and digital natives to digital wisdom. Innovate 5(3), Available at: http://www.innovateonline.info/index.php?view¼article&id¼705

Raghupathi, W., & Raghupathi, V. Big data analytics in healthcare: promise and potential. Health Information Science and Systems, (2014)

Shaw, J., (2014) Why Big Data is a Big Deal, Harvard Magazine March-April 2014, available at: http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal

ENDNOTES

[1] Anderson, C (2008) "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", WIRED, June 23 2008, www.wired.com/2008/06/pb-theory/

[2] Ibid.,

[3] Department of Electronics and Information Technology (2015) Digital India, [ONLINE] Available at: http://www.digitalindia.gov.in/. [Accessed 13 December 15].

[4] Boyd D and Crawford K (2012) Critical questions for big data. Information, Communication and Society 15(5): 662-679; Kitchen, R (2014) Big Data, new epistemologies and paradigm shifts, Big Data & Society, April-June 2014: 1-12

[5] Kuhn T (1962) The Structure of Scientific Revolutions. Chicago: University of Chicago Press.

[6] Ibid.,

[7] Hey, T., Tansley, S., and Tolle, K (eds)., (2009) The Fourth Paradigm: Data-Intensive Scientific Discovery, *Redmond: Microsoft Research*, pp. xvii-xxxi.

[8] Ibid.,

[9] Dyche J (2012) Big data 'Eurekas!' don't just happen, Harvard Business Review Blog. 20 November. Available at: http://blogs.hbr.org/cs/2012/11/eureka_doesnt_just_ happen.html

[10] Ibid.,

[11] Joh. E, (2014) 'Policing by Numbers: Big Data and the Fourth Amendment', Washington Law Review, Vol. 85: 35, https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1319/89WLR0035.pdf?sequence=1

[12] Mayer-Schonberger V and Cukier K (2013) Big Data: A Revolution that Will Change How We Live, Work and Think. London: John Murray

[13] King quoted in Shaw, J., (2014) Why Big Data is a Big Deal, *Harvard Magazine March-April 2014*, available at: http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal

[14] Boyd D and Crawford K (2012) Critical questions for big data. Information, Communication and Society 15(5): 662-679.

[15] Joh. E, 'Policing by Numbers: Big Data and the Fourth Amendment', Washington Law Review, Vol. 85: 35, (2014) https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1319/89WLR0035.pdf?sequence=1 ; Raghupathi, W., & Raghupathi, V. Big data analytics in healthcare: promise and potential. Health Information Science and Systems, (2014); Morris, D. Big data could improve supply chain efficiency-if companies would let it, Fortune, August 5 2015, http://fortune.com/2015/08/05/big-data-supply-chain/ ; , Hilbert, M. Big Data for Development: From Information- to Knowledge Societies (2013). Available at SSRN: http://ssrn.com/abstract=2205145

[16] Boyd D and Crawford K (2012) Critical questions for big data. Information, Communication and Society 15(5): 662-679; Kitchen, R (2014) Big Data, new epistemologies and paradigm shifts, Big Data & Society, April-June 2014: 1-12

[17] Prensky M (2009) H. sapiens digital: From digital immigrants and digital natives to digital wisdom. Innovate 5(3), Available at: http://www.innovateonline.info/index.php?view¼article&id¼705

[18] Hume, D., (1748), Philosophical Essays Concerning Human Understanding (1 ed.). *London: A. Millar*.

[19] Mayer-Schonberger V and Cukier K (2013) Big Data: A Revolution that Will Change How We Live, Work and Think. London: John Murray

[20] McCue, C., Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis, Butterworth-Heinemann, (2014)

[21] Kitchen, R (2014) Big Data, new epistemologies and paradigm shifts, Big Data & Society, April-June 2014: 1-12;

[22] Jasanoff, S., (2013) Watching the Watchers: Lessons from the Science of Science Advice, *Guardian 8 April 2013*, available at: http://www.theguardian.com/science/political-science/2013/apr/08/lessons-science-advice

[23] Bowker, G., (2013) The Theory-Data Thing, *International Journal of Communication* 8 (2043), 1795-1799

[24] Kitchen, R (2014) Big Data, new epistemologies and paradigm shifts, Big Data & Society, April-June 2014: 1-12

[25] Gross quoted in Ibid.,

[26] Ibid.

Benefits and Harms of Big Data

SCOTT MASON

Introduction

In 2011 it was estimated that the quantity of data produced globally would surpass 1.8 zettabyte [1]. By 2013 that had grown to 4 zettabytes [2], and with the nascent development of the so-called 'Internet of Things' gathering pace, these trends are likely to continue. This expansion in the volume, velocity, and variety of data available [3], together with the development of innovative forms of statistical analytics, is generally referred to as "Big Data"; though there is no single agreed upon definition of the term. Although still in its initial stages, Big Data promises to provide new insights and solutions across a wide range of sectors, many of which would have been unimaginable even 10 years ago.

Despite enormous optimism about the scope and variety of Big Data's potential applications however, many remain concerned about its widespread adoption, with some scholars suggesting it could generate as many harms as benefits [4]. Most notably these have included concerns about the inevitable threats to privacy associated with the generation, collection and use of large quantities of data [5]. However, concerns have also been raised regarding, for example, the lack of transparency around the design of algorithms used to process the data, over-reliance on Big Data analytics as opposed to traditional forms of analysis and the creation of new digital divides to just name a few.

The existing literature on Big Data is vast, however many of the benefits and harms identified by researchers tend to relate to sector specific applications of Big Data analytics, such as predictive policing, or targeted marketing. Whilst these examples can be useful in demonstrating the diversity of Big Data's possible applications, it can nevertheless be difficult to gain an overall perspective of the broader impacts of Big Data as a whole. As such this article will seek to disaggregate the potential benefits and harms of Big Data, organising them into several broad categories, which are reflective of the existing scholarly literature.

What are the Potential Benefits of Big Data?

From politicians to business leaders, recent years have seen Big Data confidently proclaimed as a potential solution to a diverse range of problems from, world hunger and diseases, to government budget deficits and corruption. But if we look beyond the hyperbole and headlines, what do we really know about the advantages of Big Data? Given the current buzz surrounding it, the existing literature on Big Data is perhaps unsurprisingly vast, providing innumerable examples of the potential applications of Big Data from agriculture to policing. However, rather than try (and fail) to list the many possible applications of Big Data analytics across all sectors and industries, for the purposes of this article we have instead attempted to distil the various advantages of Big Data discussed within literature into the following five broad categories; Decision-Making, Efficiency & Productivity, Research & Development, Personalisation and Transparency, each of which will be discussed separately below.

Decision-Making

Whilst data analytics have always been used to improve the quality and efficiency of decision-making processes, the advent of Big Data means that the areas of our lives in which data driven decision- making plays a role is expanding dramatically; as businesses and governments become better able to exploit new data flows. Furthermore, the real-time and predictive nature of decision-making made possible by Big Data, are increasingly allowing these decisions to be automated. As a result, Big Data is providing governments and business with unprecedented opportunities to create new insights and solutions; becoming

more responsive to new opportunities and better able to act quickly - and in some cases preemptively - to deal with emerging threats.

This ability of Big Data to speed up and improve decision-making processes can be applied across all sectors from transport to healthcare and is often cited within the literature as one of the key advantages of Big Data. Joh, for example, highlights the increased use of data driven predictive analysis by police forces to help them to forecast the times and geographical locations in which crimes are most likely to occur. This allows the force to redistribute their officers and resources according to anticipated need, and in certain cities has been highly effective in reducing crime rates [6]. Raghupathi meanwhile cites the case of healthcare, where predictive modelling driven by big data is being used to proactively identify patients who could benefit from preventative care or lifestyle changes [7].

One area in particular where the decision-making capabilities of Big Data are having a significant impact is in the field of risk management [8]. For instance, Big Data can allow companies to map their entire data landscape to help detect sensitive information, such as 16 digit numbers - potentially credit card data - which are not being stored according to regulatory requirements and intervene accordingly. Similarly, detailed analysis of data held about suppliers and customers can help companies to identify those in financial trouble, allowing them to act quickly to minimize their exposure to any potential default [9].

Efficiency and Productivity

In an era when many governments and businesses are facing enormous pressures on their budgets, the desire to reduce waste and inefficiency has never been greater. By providing the information and analysis needed for organisations to better manage and coordinate their operations, Big Data can help to alleviate such problems, leading to the better utilization of scarce resources and a more productive workforce [10].

Within the literature such efficiency savings are most commonly discussed in relation to reductions in energy consumption [11]. For example, a report published by Cisco notes how the city of Olso has managed to reduce the energy consumption of street-lighting by 62 percent through the use of smart solutions driven by Big Data [12]. Increasingly, however, statistical models generated by Big Data analytics are also being utilized to identify potential efficiencies in sourcing, scheduling and routing in a wide range of sectors from agriculture to transport. For example, Newell observes how many local governments are generating large databases of scanned license plates through the use of automated license plate recognition systems (ALPR), which government agencies can then use to help improve local traffic management and ease congestion [13].

Commonly these efficiency savings are only made possible by the often counter-intuitive insights generated by the Big Data models. For example, whilst a human analyst planning a truck route would always tend to avoid ‘drive-bys’ - bypassing one stop to reach a third before doubling back - Big Data insights can sometimes show such routes to be more efficient. In such cases efficiency saving of this kind would in all likelihood have gone unrecognised by a human analyst, not trained to look for such patterns [14].

Research, Development, and Innovation

Perhaps one of the most intriguing benefits of Big Data is its potential use in the research and development of new products and services. As is highlighted throughout the literature, Big Data can help businesses to gain an understanding of how others perceive their products or identify customer demand and adapt their marketing or indeed the design of their products accordingly [15]. Analysis of social media data, for instance, can provide valuable insights into customers’ sentiments towards existing products as well as discover demands for new products and services, allowing businesses to respond more quickly to changes in customer behaviour [16].

In addition to market research, Big Data can also be used during the design and development stage of new products; for example by helping to test thousands of different variations of computer-aided designs in an expedient and cost-effective manner. In doing so, business and designers are able to better assess how minor changes to a products design may affect its cost and performance, thereby improving the cost-effectiveness of the production process and increasing profitability.

Personalisation

For many consumers, perhaps the most familiar application of Big Data is its ability to help tailor products and services to meet their individual preferences. This phenomena is most immediately noticeable on many online services such as Netflix; where data about users activities and preferences is collated and analysed to provide a personalised service, for example by suggesting films or television shows the user may enjoy based upon their previous viewing history [17]. By enabling companies to generate in-depth profiles of their customers, Big Data allows businesses to move past the 'one size fits all' approach to product and services design and instead quickly and cost-effectively adapt their services to better meet customer demand.

In addition to service personalisation, similar profiling techniques are increasingly being utilized in sectors such as healthcare. Here data about a patient's medical history, lifestyle, and even their gene expression patterns are collated, generating a detailed medical profile which can then be used to tailor treatments to meet their specific needs [18]. Targeted care of this sort can not only help to reduce costs for example by helping to avoid over-prescriptions, but may also help to improve the effectiveness of treatments and so ultimately their outcome.

Transparency

If 'knowledge is power', then, - so say Big Data enthusiasts - advances in data analytics and the quantity of data available can give consumers and citizens the knowledge to hold governments and businesses to account, as well as make more informed choices about the products and services they use. Nevertheless, data (even lots of it) does not necessarily equal knowledge. In order for citizens and consumers to be able to fully utilize the vast quantities of data available to them, they must first have some way to make sense of it. For some, Big Data analytics provides just such a solution, allowing users to easily search, compare and analyze available data, thereby helping to challenge existing information asymmetries and make business and government more transparent [19].

In the private sector, Big Data enthusiasts have claimed that Big Data holds the potential to ensure complete transparency of supply chains, enabling concerned consumers to trace the source of their products, for example to ensure that they have been sourced ethically [20]. Furthermore, Big Data is now making accessible information which was previously unavailable to average consumers and challenging companies whose business models rely on the maintenance of information asymmetries.The real-estate industry, for example, relies heavily upon its ability to acquire and control proprietary information, such as transaction data as a competitive asset. In recent years, however, many online services have allowed consumers to effectively bypass agents, by providing alternative sources of real-estate data and enabling prospective buyers and sellers to communicate directly with each other [21]. Therefore, providing consumers with access to large quantities of actionable data . Big Data can help to eliminate established information asymmetries, allowing them to make better and more informed decisions about the products they buy and the services they enlist.

This potential to harness the power of Big Data to improve transparency and accountability can also be seen in the public sector, with many scholars suggesting that greater access to government data could help to stem corruption and make politics more accountable. This view was recently endorsed by the UN who highlighted the potential uses of Big Data to improve policymaking and accountability in a report published by the Independent Expert

Advisory Group on the "Data Revolution for Sustainable Development". In the report experts emphasize the potential of what they term the 'data revolution', to help achieve sustainable development goals by for example helping civil society groups and individuals to 'develop data literacy and help communities and individuals to generate and use data, to ensure accountability and make better decisions for themselves' [22].

What are the Potential Harms of Big Data?

Whilst it is often easy to be seduced by the utopian visions of Big Data evangelists, in order to ensure that Big Data can deliver the types of far-reaching benefits its proponents promise, it is vital that we are also sensitive to its potential harms. Within the existing literature, discussions about the potential harms of Big Data are perhaps understandably dominated by concerns about privacy. Yet as Big Data has begun to play an increasingly central role in our daily lives, a broad range of new threats have begun to emerge including issues related to security and scientific epistemology, as well as problems of marginalisation, discrimination and transparency; each of which will be discussed separately below.

Privacy

By far the biggest concern raised by researchers in relation to Big Data is its risk to privacy. Given that by its very nature Big Data requires extensive and unprecedented access to large quantities of data; it is hardly surprising that many of the benefits outlined above in one way or another exist in tension with considerations of privacy. Although many scholars have called for a broader debate on the effects of Big Data on ethical best practice [23], a comprehensive exploration into the complex debates surrounding the ethical implications of Big Data go far beyond the scope of this article. Instead we will simply attempt to highlight some of the major areas of concern expressed in the literature, including its effects on established principles of privacy and the implication of Big Data on the suitability of existing regulatory frameworks governing privacy and data protection.

Re-identification

Traditionally many Big Data enthusiasts have used de-identification - the process of anonymising data by removing personally identifiable information (PII) - as a way of justifying mass collection and use of personal data. By claiming that such measures are sufficient to ensure the privacy of users, data brokers, companies and governments have sought to deflect concerns about the privacy implications of Big Data, and suggest that it can be compliant with existing regulatory and legal frameworks on data protection.

However, many scholars remain concerned about the limits of anonymisation. As Tene and Polonetsky observe 'Once data-such as a clickstream or a cookie number-are linked to an identified individual, they become difficult to disentangle' [24]. They cite the example of University of Texas researchers Narayanan and Shmatikov, who were able to successfully re-identify anonymised Netflix user data by cross referencing it with data stored in a publicly accessible online database. As Narayanan and Shmatikov themselves explained, 'once any piece of data has been linked to a person's real identity, any association between this data and a virtual identity breaks anonymity of the latter' [25]. The quantity and variety of datasets which Big Data analytics has made associable with individuals is therefore expanding the scope of the types of data that can be considered PII, as well as undermining claims that de-identification alone is sufficient to ensure privacy for users.

Privacy Frameworks Obsolete?

In recent decades privacy and data protection frameworks based upon a number of so-called 'privacy principles' have formed the basis of most attempts to encourage greater consideration of privacy issues online [26]. For many however, the emergence of Big Data has raised question about the extent to which these 'principles of privacy' are workable in an era of ubiquitous data collection.

Collection Limitation and Data Minimization: Big Data by its very nature requires the collection and processing of very large and very diverse data sets. Unlike other forms scientific research and analysis which utilize various sampling techniques to identify and target the types of data most useful to the research questions, Big Data instead seeks to gather as much data as possible, in order to achieve full resolution of the phenomenon being studied, a task made much easier in recent years as a result of the proliferation of internet enabled devices and the growth of the Internet of Things. This goal of attaining comprehensive coverage exists in tension however with the key privacy principles of collection limitation and data minimization which seek to limit both the quantity and variety of data collected about an individual to the absolute minimum [27].

Purpose Limitation: Since the utility of a given dataset is often not easily identifiable at the time of collection, datasets are increasingly being processed several times for a variety of different purposes. Such practices have significant implications for the principle of purpose limitation, which aims to ensure that organizations are open about their reasons for collecting data, and that they use and process the data for no other purpose than those initially specified [28].

Notice and Consent: The principles of notice and consent have formed the cornerstones of attempts to protect privacy for decades. Nevertheless in an era of ubiquitous data collection, the notion that an individual must be required to provide their explicit consent to allow for the collection and processing of their data seems increasingly antiquated, a relic of an age when it was possible to keep track of your personal data relationships and transactions. Today as data streams become more complex, some have begun to question suitability of consent as a mechanism to protect privacy. In particular commentators have noted how given the complexity of data flows in the digital ecosystem most individuals are not well placed to make truly informed decisions about the management of their data [29]. In one study, researchers demonstrated how by creating the perceptions of control, users were more likely to share their personal information, regardless of whether or not the users had actually gained control [30]. As such, for many, the garnering of consent is increasingly becoming a symbolic box-ticking exercise which achieves little more than to irritate and inconvenience customers whilst providing a burden for companies and a hindrance to growth and innovation [31].

Access and Correction: The principle of 'access and correction' refers to the rights of individuals to obtain personal information being held about them as well as the right to erase, rectify, complete or otherwise amend that data. Aside from the well documented problems with privacy self-management, for many the real-time nature of data generation and analysis in an era of Big Data poses a number of structural challenges to this principle of privacy. As x comments, 'a good amount of data is not pre-processed in a similar fashion as traditional data warehouses. This creates a number of potential compliance problems such as difficulty erasing, retrieving or correcting data. A typical big data system is not built for interactivity, but for batch processing. This also makes the application of changes on a (presumably) static data set difficult' [32].

Opt In-Out: The notion that the provision of data should be a matter of personal choice on the part of the individual and that the individual can, if they chose decide to 'opt-out' of data collection, for example by ceasing use of a particular service, is an important component of privacy and data protection frameworks. The proliferation of internet-enabled devices, their integration into the built environment and the real-time nature of data collection and analysis however are beginning to undermine this concept. For many critics of Big Data the ubiquity of data collection points as well as the compulsory provision of data as a prerequisite for the access and use of many key online services is making opting-out of data collection not only impractical but in some cases impossible [33].

"Chilling Effects"

For many scholars the normalization of large scale data collection is steadily producing a widespread perception of ubiquitous surveillance amongst users. Drawing upon Foucault's analysis of Jeremy Bentham's panopticon and the disciplinary effects of surveillance, they argue that this perception of permanent visibility can cause users to sub-consciously 'discipline' and self- regulate of their own behavior, fearful of being targeted or identified as 'abnormal' [34]. As a result, the pervasive nature of Big Data risks generating a 'chilling effect' on user behavior and free speech.

Although the notion of "chilling effects" is quite prevalent throughout the academic literature on surveillance and security, the difficulty of quantifying the perception and effects of surveillance on online behavior and practices means that there have only been a limited number of empirical studies of this phenomena, and none directly related to the chilling effects of Big Data. One study, conducted by researchers at MIT however, sought to assess the impact of Edward Snowden's revelations about NSA surveillance programs on Google search trends. Nearly 6,000 participants were asked to individually rate certain keywords for their perceived degree of privacy sensitivity along multiple dimensions. Using Google's own publicly available search data, the researchers then analyzed search patterns for these terms before and after the Snowden revelations. In doing so they were able to demonstrate a reduction of around 2.2% in searchers for those terms deemed to be most sensitive in nature. According to the researchers themselves, the results 'suggest that there is a chilling effect on search behaviour from government surveillance on the Internet' [35]. Although this study focussed on the effects on government surveillance, for many privacy advocates the growing pervasiveness of Big Data risks generating similar results [36].

Dignitary Harms of Predictive Decision-Making

In addition to its potentially chilling effects on free speech, the automated nature of Big Data analytics also possess the potential to inflict so-called 'dignitary harms' on individuals, by revealing insights about themselves that they would have preferred to keep private [37].

In an infamous example, following a shopping trip to the retail chain Target, a young girl began to receive mail at her father's house advertising products for babies including, diapers, clothing, and cribs. In response, her father complained to the management of the company, incensed by what he perceived to be the company's attempts to "encourage" pregnancy in teens. A few days later however, the father was forced to contact the store again to apologies, after his daughter had confessed to him that she was indeed pregnant. It was later revealed that Target regularly analyzed the sale of key products such as supplements or unscented lotions in order to generate "pregnancy prediction" scores, which could be used to assess the likelihood that a customer was pregnant and to therefore target them with relevant offers [38]. Such cases, though anecdotal illustrate how Big Data if not adopted sensitively can lead to potential embarrassing information about users being made public.

Security

In relation to cybersecurity Big Data can be viewed to a certain extent as a double-edged sword. On the one hand, the unique capabilities of Big Data analytics can provide organizations with new and innovative methods of enhancing their cybersecurity systems. On the other however, the sheer quantity and diversity of data emanating from a variety of sources creates its own security risks.

"Honey-Pot"

The larger the quantities of confidential information stored by companies on their databases the more attractive those databases may appear to potential hackers.

Data Redundancy and Dispersion

Inherent to Big Data systems is the duplication of data to many locations in order to optimize query processing. Data is dispersed across a wide range of data repositories in different servers, in different parts of the world. As a result it may be difficult for organizations to accurately locate and secure all items of personal information.

Epistemological and Methodological Implications

In 2008 Chris Anderson infamously proclaimed the 'end of theory'. Writing for Wired Magazine, Anderson predicted that the coming age of Big Data would create a 'deluge of data' so large that the scientific methods of hypothesis, sampling and testing would be rendered 'obsolete' [39]. 'There is now a better way' Anderson insisted, 'Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot' [40].

In spite of these bold claims however, many theorists remain skeptical of Big Data's methodological benefits and have expressed concern about its potential implications for conventional scientific epistemologies. For them the increased prominence of Big Data analytics in science does not signal a paradigmatic transition to a more enlightened data-driven age, but a hollowing out of the scientific method and an abandonment of casual knowledge in favor of shallow correlative analysis [41].

Obfuscation

Although Big Data analytics can be utilized to study almost any phenomena where enough data exists, many theorists have warned that simply because Big Data analytics *can* be used does not necessarily mean that they *should* be used [42]. Bigger is not always better and indeed the sheer quantity of data made available to users may in fact act to obscure certain insights. Whereas traditional scientific methods use sampling techniques to identify the most important and relevant data, Big Data by contrast encourages the collection and use of as much data as possible, in an attempt to attain full resolution of the phenomena being studied. However, not all data is equally useful and simply inputting as much data as possible into an algorithm is unlikely to produce accurate results and may instead obscure key insights.

Indeed, whilst the promise of automation is central to a large part of Big Data's appeal, researchers observe that most Big Data analysis still requires an element of human judgement to filter out the 'good' data from the 'bad', and to decide what aspects of the data are relevant to the research objectives. As Boyd and Crawford observe, 'in the case of social media data, there is a 'data cleaning' process: making decisions about what attributes and variables will be counted, and which will be ignored. This process is inherently subjective" [43].

Google's Flu Trend project provides an illustrative example of how Big Data's tendency to try to maximise data inputs can produce misleading results. Designed to accurately track flu outbreaks based upon data collected from Google searches, the project was initially proclaimed to be a great success. Gradually however it became apparent that the results being produced were not reflective of the reality on the ground. Later it was discovered that the algorithms used by the project to interpret search terms were insufficiently accurate to filter out anomalies in searches, such as those related to the 2009 H1N1 flu pandemic. As such, despite the great promise of Big Data, scholars insist it remains critical to be mindful of its limitations, remain selective about the types of data included in the analysis and exercise caution and intuition whenever interpreting its results [44].

"Apophenia"

In complete contrast to the problem of obfuscation, Boyd and Crawford observe how Big Data may also lead to the practice of 'apophenia', a phenomena whereby analysts interpret patterns where none exist, 'simply because enormous quantities of data can offer connections that radiate in all directions" [45]. David Leinweber for example demonstrated that data mining techniques could show strong but ultimately spurious correlations between changes in the S&P 500 stock index and butter production in Bangladesh [46]. Such spurious correlation between disparate and unconnected phenomena are a common feature of Big Data analytics and risks leading to unfounded conclusions being draw from the data.

Although Leinweber's primary focus of analysis was the use of Data-Mining technologies, his observations are equally applicable to Big Data. Indeed the tendency amongst Big Data analysts to marginalise the types of domain specific expertise capable of differentiating between relevant and irrelevant correlations in favour of algorithmic automation can in many ways be seen to exacerbate many of the problems Leinweber identified.

From Causation to Correlation

Closely related to the problem of Aphonenia is the concern that Big Data's emphasis on correlative analysis risks leading to an abandonment of the pursuit of causal knowledge in favour of shallow descriptive accounts of scientific phenomena [47].

For many, Big Data enthusiasts 'correlation is enough', producing inherently meaningful results interpretable by anyone without the need for pre-existing theory or hypothesis. Whilst proponents of Big Data claim that such an approach allows them to produce objective knowledge, by cleansing the data of any kind of philosophical or ideological commitment, for others by neglecting the knowledge of domain experts, Big Data risks generating a shallow type of analysis, since it fails to adequately embed observations within a pre-existing body of knowledge.

This commitment to an empiricist epistemology and methodological monism is particularly problematic in the context of studies of human behaviour, where actions cannot be calculated and anticipated using quantifiable data alone. In such instances, a certain degree of qualitative analysis of social, historical and cultural variables may be required in order to make the data meaningful by embedding it within a broader body of knowledge. The abstract and intangible nature of these variables requires a great deal of expert knowledge and interpretive skill to comprehend. It is therefore vital that the knowledge of domain specific experts is properly utilized to help 'evaluate the inputs, guide the process, and evaluate the end products within the context of value and validity' [48].

As such, although Big Data can provide unrivalled accounts of "what" people do, it fundamentally fails to deliver robust explanations of "why" people do it. This problem is especially critical in the case of public policy-making since without any indication of the motivations of individuals, policy-makers can have no basis upon which to intervene to incentivise more positive outcomes.

Digital Divides and Marginalisation

Today data is a highly valuable commodity. The market for data in and of itself has been steadily growing in recent years with the business models of many online services now formulated around the strategy of harvesting data from users [49]. As with the commodification of anything however, inequalities can easily emerge between the haves and have not's. Whilst the quantity of data currently generated on a daily basis is many times greater than at any other point in human history, the vast majority of this data is owned and tightly controlled by a very small number of technology companies and data brokers. Although in some instances limited access to data may be granted to university researchers or to those willing and able to pay a fee, in many cases data remains jealously guarded by data brokers, who view it as an important competitive asset. As a result these data brokers

and companies risk becoming the gatekeepers of the Big Data revolution, adjudicating not only over who can benefit from Big Data, but also in what context and under what terms. For many such inconsistencies and inequalities in access to data raises serious doubts about just how widely distributed the benefits of Big Data will be. Others go even further claiming that far from helping to alleviate inequalities, the advent of Big Data risks exacerbating already significant digital divides that exist as well as creating new ones [50].

Anti-Competitive Practices

As a result of the reluctance of large companies to share their data, there increasingly exists a divide in access between small start-ups companies and their larger and more established competitors. Thus, new entrants to the marketplace may be at a competitive disadvantage in relation to large and well established enterprises, being as they are unable to harness the analytical power of the vast quantities of data available to large companies by virtue of their privileged market position. Since the performance of many online services are today often intimately connected with the collation and use of users data, some researchers have suggested that this inequity in access to data could lead to a reduction in competition in the online marketplace, and ultimately therefore to less innovation and choice for consumers [51].

As a result researchers including Nathan Newman of New York University have called for a reassessment and reorientation of anti-trust investigations and regulatory approaches more generally to 'to focus on how control of personal data by corporations can entrench monopoly power and harm consumer welfare in an economy shaped increasingly by the power of "big data"' [52]. Similarly a report produced by the European Data Protection Supervisor concluded that, 'The scope for abuse of market dominance and harm to the consumer through refusal of access to personal information and opaque or misleading privacy policies may justify a new concept of consumer harm for competition enforcement in digital economy' [53].

Research

From a research perspective barriers to access to data caused by proprietary control of datasets are problematic, since certain types of research could become restricted to those privileged enough to be granted access to data. Meanwhile those denied access are left not only incapable of conducting similar research projects, but also unable to test, verify or reproduce the findings of those who do. The existence of such gatekeepers may also lead to reluctance on the part of researchers to undertake research critical of the companies, upon whom they rely for access, leading to a chilling effect on the types of research conducted [54].

Inequality

Whilst bold claims are regularly made about the potential of Big Data to deliver economic development and generate new innovations, some critics of remain concerned about how equally the benefits of Big Data will be distributed and the effects this could have on already established digital divides [55].

Firstly, whilst the power of Big Data is already being utilized effectively by most economically developed nations, the same cannot necessarily be said for many developing countries. A combination of lower levels of connectivity, poor information infrastructure, underinvestment in information technologies and a lack of skills and trained personnel make it far more difficult for the developing world to fully reap the rewards of Big Data. As a consequence the Big Data revolution risks deepening global economic inequality as developing countries find themselves unable to compete with data rich nations whose governments can more easily exploit the vast quantities of information generated by their technically literate and connected citizens.

Likewise, to the extent that the Big Data analytics is playing a greater role in public policy-making, the capacity of individuals to generate large quantities of data, could potentially impact upon the extent to which they can provide inputs into the policy-making process. In a country such as India for example, where there exist high levels of inequality in access to information and communication technologies and the internet, there remain large discrepancies in the quantities of data produced by individuals. As a result there is a risk that those who lack access to the means of producing data will be disenfranchised, as policy-making processes become configured to accommodate the needs and interests of a privilege minority [56].

Discrimination

Injudicious or Discriminatory Outcomes

Big Data presents the opportunity for governments, businesses and individuals to make better, more informed decisions at a much faster pace. Whilst this can evidently provide innumerable opportunities to increase efficiency and mitigate risk, by removing human intervention and oversight from the decision-making process Big Data analysts run the risk of becoming blind to unfair or injudicious results generated by skewed or discriminatory programming of the algorithms.

There currently exists a large number of automated decision-making algorithms in operation across a broad range of sectors including most notably perhaps those used to assess an individual's suitability for insurance or credit. In either of these cases faults in the programming or discriminatory assessment criteria can have potentially damaging implications for the individual, who may as a result be unable to attain credit or insurance. This concern with the potentially discriminatory aspects of Big Data is prevalent throughout the literature and real life examples have been identified by researchers in a large number of major sectors in which Big Data is currently being used [57].

Yu for instance, cites the case of the insurance company Progressive, which required its customers to install 'Snapsnot' - a small monitoring device - into their cars in order to receive their best rates. The device tracked and reported the customers driving habits, and offered discounts to those drivers who drove infrequently, broke smoothly, and avoided driving at night - behaviors that correlate with a lower risk of future accidents. Although this form of price differentiation provided incentives for customers to drive more carefully, it also had the unintended consequence of unfairly penalizing late-night shift workers. As Yu observes, 'for late night shift-workers, who are disproportionately poorer and from minority groups, this differential pricing provides no benefit at all. It categorizes them as similar to late-night party-goers, forcing them to carry more of the cost of the intoxicated and other irresponsible driving that happens disproportionately at night' [58].

In another example, it is noted how Big Data is increasingly being used to evaluate applicants for entry-level service jobs. One method of evaluating applicants is by the length of their commute - the rationale being that employees with shorter commutes are statistically more likely to remain in the job longer. However, since most service jobs are typically located in town centers and since poorer neighborhoods tend to be those on the outskirts of town, such criteria can have the effect of unfairly disadvantaging those living in economically deprived areas. Consequently such metrics of evaluation can therefore also unintentionally act to reinforce existing social inequalities by making it more difficult for economically disadvantaged communities to work their way out of poverty [59].

Lack of Algorithmic Transparency

If data is indeed the 'oil of the 21st century' [60] then algorithms are very much the engines which are driving innovation and economic development. For many companies the quality of their algorithms is often a crucial factor in providing them with a market advantage over their competitor. Given their importance, the secrets behind the programming of algorithms are

often closely guarded by companies, and are typically classified as trade secrets and as such are protected by intellectual property rights. Whilst companies may claim that such secrecy is necessary to encourage market competition and innovation, many scholars are becoming increasingly concerned about the lack of transparency surrounding the design of these most crucial tools.

In particular there is a growing sentiment common amongst many researchers that there currently exists a chronic lack of accountability and transparency in terms of how Big Data algorithms are programmed and what criteria are used to determine outcomes [61]. As Frank Pasquale observed,

'hidden algorithms can make (or ruin) reputations, decide the destiny of entrepreneurs, or even devastate an entire economy. Shrouded in secrecy and complexity, decisions at major Silicon Valley and Wall Street firms were long assumed to be neutral and technical. But leaks, whistleblowers, and legal disputes have shed new light on automated judgment. Self-serving and reckless behavior is surprisingly common, and easy to hide in code protected by legal and real secrecy' [62].

As such, without increased transparency in algorithmic design, instances of Big Data discrimination may go unnoticed as analyst are unable to access the information necessary to identify them.

Conclusion

Today Big Data presents us with as many challenges as it does benefits. Whilst Big Data analytics can offer incredible opportunities to reduce inefficiency, improve decision-making, and increase transparency, concerns remain about the effects of these new technologies on issues such as privacy, equality and discrimination. Although the tensions between the competing demands of Big Data advocates and their critics may appear irreconcilable; only by highlighting these points of contestation can we hope to begin to ask the types of important and difficult questions necessary to do so, including; how can we reconcile Big Data's need for massive inputs of personal information with core principles of privacy such as data minimization and collection limitation? What processes and procedures need to be put in place during the design and implementation of Big Data models and algorithms to provide sufficient transparency and accountability so as to avoid instances of discrimination? What measures can be used to help close digital divides and ensure that the benefits of Big Data are shared equitably? Questions such as these are today only just beginning to be addressed; each however, will require careful consideration and reasoned debate, if Big Data is to deliver on its promises and truly fulfil its 'revolutionary' potential.

ENDNOTES

[1] Gantz, J., &Reinsel, D. Extracting Value from Chaos, *IDC*, (2011), available at:http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

[2] Meeker, M. & Yu, L. Internet Trends, *Kleiner Perkins Caulfield Byers*, (2013),http://www.slideshare.net/kleinerperkins/kpcb-internet-trends-2013 .

[3] Douglas, L. "*3D Data Management: Controlling Data Volume, Velocity and Variety*". *Gartner*, (2001)

[4] Boyd, D., and Crawford, K. 'Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, Vol 15, Issue 5, (2012) http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878, Tene, O., &Polonetsky, J. Big Data for All: Privacy and User Control in the Age of Analytics, *11 Nw. J. Tech. &Intell. Prop. 239* (2013) http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

[5] Ibid.

[6] Joh. E, 'Policing by Numbers: Big Data and the Fourth Amendment', *Washington Law Review*, Vol. 85: 35, (2014)https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1319/89WLR0035.pdf?sequence=1

[7] Raghupathi, W., &Raghupathi, V. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, (2014)

[8] Anderson, R., & Roberts, D. 'Big Data: Strategic Risks and Opportunities, *Crowe Horwarth Global Risk Consulting Limited*, (2012) https://www.crowehorwath.net/uploadedfiles/crowe-horwath-global/tabbed_content/big%20data%20strategic%20risks%20and%20opportunities%20white%20paper_risk13905.pdf

[9] Ibid.

[10] Kshetri. N, 'The Emerging role of Big Data in Key development issues: Opportunities, challenges, and concerns'. *Big Data & Society* (2014) http://bds.sagepub.com/content/1/2/2053951714564227.abstract,

[11] Tene, O., &Polonetsky, J. Big Data for All: Privacy and User Control in the Age of Analytics, *11 Nw. J. Tech. &Intell. Prop. 239* (2013) http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

[12] Cisco, 'IoE-Driven Smart Street Lighting Project Allows Oslo to Reduce Costs, Save Energy, Provide Better Service', Cisco, (2014) Available at: http://www.cisco.com/c/dam/m/en_us/ioe/public_sector/pdfs/jurisdictions/Oslo_Jurisdiction_Profile_051214REV.pdf

[13] Newell, B, C. Local Law Enforcement Jumps on the Big Data Bandwagon: Automated License Plate Recognition Systems, Information Privacy, and Access to Government Information. *University of Washington - the Information School*, (2013) http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2341182

[14] Morris, D. Big data could improve supply chain efficiency-if companies would let it, *Fortune*, August 5 2015, http://fortune.com/2015/08/05/big-data-supply-chain/

[15] Tucker, Darren S., & Wellford, Hill B., Big Mistakes Regarding Big Data, Antitrust Source, American Bar Association, (2014). Available at SSRN: http://ssrn.com/abstract=2549044

[16] Davenport, T., Barth., Bean, R. How is Big Data Different, *MITSloan Management Review*, Fall (2012), Available at, http://sloanreview.mit.edu/article/how-big-data-is-different/

[17] Tucker, Darren S., & Wellford, Hill B., Big Mistakes Regarding Big Data, Antitrust Source, American Bar Association, (2014). Available at SSRN: http://ssrn.com/abstract=2549044

[18] Raghupathi, W., &Raghupathi, V. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, (2014)

[19] Brown, B., Chui, M., Manyika, J. 'Are you Ready for the Era of Big Data?', *McKinsey Quarterly*, (2011), Available at, http://www.t-systems.com/solutions/download-mckinsey-quarterly-/1148544_1/blobBinary/Study-McKinsey-Big-data.pdf ; Benady, D., 'Radical transparency will be unlocked by technology and big data', *Guardian* (2014) Available at: http://www.theguardian.com/sustainable-business/radical-transparency-unlocked-technology-big-data

[20] Ibid.

[21] Ibid.

[22] United Nations, A World That Counts: Mobilising the Data Revolution for Sustainable Development, *Report prepared at the request of the United Nations Secretary-General,by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development*. (2014), pg. 18, see also, Hilbert, M. Big Data for Development: From Information- to Knowledge Societies (2013). Available at SSRN: http://ssrn.com/abstract=2205145

[23] Greenleaf, G. Abandon All Hope? *Foreword for Issue 37(2) of the UNSW Law Journal on 'Communications Surveillance, Big Data, and the Law*' ,(2014) http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2490425##, Boyd, D., and Crawford, K. 'Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, Vol. 15, Issue 5, (2012) http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878

[24] Tene, O., & Polonetsky, J. Big Data for All: Privacy and User Control in the Age of Analytics, *11 Nw. J. Tech. & Intell. Prop. 239* (2013) http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

[25] Narayanan and Shmatikov quoted in Ibid.,

[26] OECD, Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, The Organization for Economic Co-Operation and Development, (1999); The European Parliament and the Council of the European Union, EU Data Protection Directive, "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," (1995)

[27] Barocas, S., & Selbst, A, D., Big Data's Disparate Impact, *California Law Review, Vol. 104*, (2015). Available at SSRN: http://ssrn.com/abstract=2477899

[28] Article 29 Working Group., Opinion 03/2013 on purpose limitation, *Article 29 Data Protection Working Party*, (2013) available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

[29] Solove, D, J. Privacy Self-Management and the Consent Dilemma, 126 Harv. L. Rev. 1880 (2013), Available at: http://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2093&context=faculty_publications

[30] Brandimarte, L., Acquisti, A., & Loewenstein, G., Misplaced Confidences: Privacy and the Control Paradox, *Ninth Annual Workshop on the Economics of Information Security (WEIS) June 7-8 2010, Harvard University, Cambridge, MA,* (2010), available at: https://fpf.org/wp-content/uploads/2010/07/Misplaced-Confidences-acquisti-FPF.pdf

[31] Solove, D, J., Privacy Self-Management and the Consent Dilemma, *126 Harv. L. Rev. 1880* (2013), Available at: http://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2093&context=faculty_publications

[32] Yu, W, E., Data., Privacy and Big Data-Compliance Issues and Considerations, *ISACA Journal, Vol. 3 2014* (2014), available at: http://www.isaca.org/Journal/archives/2014/Volume-3/Pages/Data-Privacy-and-Big-Data-Compliance-Issues-and-Considerations.aspx

[33] Ramirez, E., Brill, J., Ohlhausen, M., Wright, J., & McSweeny, T., Data Brokers: A Call for Transparency and Accountability, *Federal Trade Commission* (2014) https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

[34] Michel Foucault, Discipline and Punish: The Birth of the Prison. Translated by Alan Sheridan, *London: Allen Lane, Penguin*, (1977)

[35] Marthews, A., & Tucker, C., Government Surveillance and Internet Search Behavior (2015), available at SSRN: http://ssrn.com/abstract=2412564

[36] Boyd, D., and Crawford, K. 'Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, Vol. 15, Issue 5, (2012)

[37] Hirsch, D., That's Unfair! Or is it? Big Data, Discrimination and the FTC's Unfairness Authority, *Kentucky Law Journal, Vol. 103*, available at: http://www.kentuckylawjournal.org/wp-content/uploads/2015/02/103KyLJ345.pdf

[38] Hill, K., How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

[39] Anderson, C (2008) "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", WIRED, June 23 2008, www.wired.com/2008/06/pb-theory/

[40] Ibid.,

[41] Kitchen, R (2014) Big Data, new epistemologies and paradigm shifts, Big Data & Society, April-June 2014: 1-12

[42] Boyd D and Crawford K (2012) Critical questions for big data. Information, Communication and Society 15(5): 662-679

[43] Ibid

[44] Lazer, D., Kennedy, R., King, G., &Vespignani, A. " The Parable of Google Flu: Traps in Big Data Analysis ." *Science 343* (2014): 1203-1205. Copy at http://j.mp/1ii4ETo

[45] Boyd, D., and Crawford, K. 'Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, Vol 15, Issue 5, (2012) http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878

[46] Leinweber, D. (2007) 'Stupid data miner tricks: overfitting the S&P 500', The Journal of Investing, vol. 16, no. 1, pp. 15-22. http://m.shookrun.com/documents/stupidmining.pdf

[47] Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication and Society* 15(5): 662-679

[48] McCue, C., Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis, *Butterworth-Heinemann*, (2014)

[49] De Zwart, M. J., Humphreys, S., & Van Dissel, B. Surveillance, big data and democracy: lessons for Australia from the US and UK. *Http://www.unswlawjournal.unsw.edu.au/issue/volume-37-No-2.* (2014) Retrieved from https://digital.library.adelaide.edu.au/dspace/handle/2440/90048

[50] Boyd, D., and Crawford, K. 'Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, Vol 15, Issue 5, (2012) http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878; Newman, N., Search, Antitrust and the Economics of the Control of User Data, 31 YALE J. ON REG. 401 (2014)

[51] Newman, N., The Cost of Lost Privacy: Search, Antitrust and the Economics of the Control of User Data (2013). Available at SSRN: http://ssrn.com/abstract=2265026, Newman, N., Search, Antitrust and the Economics of the Control of User Data, 31 YALE J. ON REG. 401 (2014)

[52] Ibid.,

[53] European Data Protection Supervisor, Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy, (2014), available at: https://secure.edps.europa.eu/EDPSWEB/webdav/shared/Documents/Consultation/Opinions/2014/14-03-26_competitition_law_big_data_EN.pdf

[54] Boyd, D., and Crawford, K. 'Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, Vol 15, Issue 5, (2012) http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878

[55] Schradie, J., Big Data Not Big Enough? How the Digital Divide Leaves People Out, MediaShift, 31 July 2013, (2013), available at: http://mediashift.org/2013/07/big-data-not-big-enough-how-digital-divide-leaves-people-out/

[56] Crawford, K., The Hidden Biases in Big Data, *Harvard Business Review, 1 April 2013* (2013), available at: https://hbr.org/2013/04/the-hidden-biases-in-big-data

[57] Robinson, D., Yu, H., Civil Rights, Big Data, and Our Algorithmic Future, (2014) http://bigdata.fairness.io/introduction/

[58] Ibid.

[59] Ibid

[60] Rotellla, P., Is Data The New Oil? Forbes, 2 April 2012, (2012), available at: http://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/

[61] Barocas, S., &Selbst, A, D., Big Data's Disparate Impact, *California Law Review, Vol. 104*, (2015). Available at SSRN: http://ssrn.com/abstract=2477899; Kshetri. N, 'The Emerging role of Big Data in Key development issues: Opportunities, challenges, and concerns'. *Big Data & Society* (2014) http://bds.sagepub.com/content/1/2/2053951714564227.abstract

[62] Pasquale, F., The Black Box Society: The Secret Algorithms That Control Money and Information, Harvard University Press, (2015)

A Review of the Policy Debate around Big Data and Internet of Things

ELONNAI HICKOK

Defining and Connecting Big Data and Internet of Things

The Internet of Things is a term that refers to networked objects and systems that can connect to the internet and can transmit and receive data. Characteristics of IoT include the gathering of information through sensors, the automation of functions, and analysis of collected data [1]. For IoT devices, because of the *velocity* at which data is generated, the *volume* of data that is generated, and the *variety* of data generated by different sources [2] - IoT devices can be understood as generating Big Data and/or relying on Big Data analytics. In this way IoT devices and Big Data are intrinsically interconnected.

General Implications of Big Data and Internet of Things

Big Data paradigms are being adopted across countries, governments, and business sectors because of the potential insights and change that it can bring. From improving an organizations business model, facilitating urban development, allowing for targeted and individualized services, and enabling the prediction of certain events or actions - the application of Big Data has been recognized as having the potential to bring about dramatic and large scale changes.

At the same time, experts have identified risks to the individual that can be associated with the generation, analysis, and use of Big Data. In May 2014, the White House of the United States completed a ninety day study of how big data will change everyday life. The Report highlights the potential of Big Data as well as identifying a number of concerns associated with Big Data. For example: the selling of personal data, identification or re-identification of individuals, profiling of individuals, creation and exacerbation of information asymmetries, unfair, discriminating, biased, and incorrect decisions based on Big Data analytics, and lack of or misinformed user consent [3]. Errors in Big Data analytics that experts have identified include statistical fallacies, human bias, translation errors, and data errors [4]. Experts have also discussed fundamental changes that Big Data can bring about. For example, Danah Boyd and Kate Crawford in the article "*Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon*" propose that Big Data can change the definition of knowledge and shape the reality it measures [5]. Similarly, a BSC/Oxford Internet Institute conference report titled "*The Societal Impact of the Internet of Things*" points out that often users of Big Data assume that information and conclusions based on digital data is reliable and in turn replace other forms of information with digital data [6].

Concerns that have been voiced by the Article 29 Working Party and others specifically about IoT devices have included insufficient security features built into devices such as encryption, the reliance of the devices on wireless communications, data loss from infection by malware or hacking, unauthorized access and use of personal data, function creep resulting from multiple IoT devices being used together, and unlawful surveillance [7].

Regulation of Big Data and Internet of Things

The regulation of Big Data and IoT is currently being debated in contexts such as the US and the EU. Academics, civil society, and regulators are exploring questions around the adequacy of present regulation and overseeing frameworks to address changes brought about Big Data, and if not - what forms of or changes in regulation are needed? For example, Kate Crawford and Jason Shultz in the article "*Big Data and Due Process: Towards a Framework to Redress*

Predictive Privacy Harms"stress the importance of bringing in 'data due process rights' i.e ensuring fairness in the analytics of Big Data and how personal information is used [8]. While Solon Barocas and Andrew Selbst in the article "*Big Data's Disparate Impact*" explore if present anti-discrimination legislation and jurisprudence in the US is adequate to protect against discrimination arising from Big Data practices - specifically data mining [9].

The Impact of Big Data and IoT on Data Protection Principles

In the context of data protection, various government bodies, including the Article 29 Data Protection Working Party set up under the Directive 95/46/EC of the European Parliament, the Council of Europe, the European Commission, and the Federal Trade Commission, as well as experts and academics in the field, have called out at least ten different data protection principles and concepts that Big Data impacts:

1. **Collection Limitation**: As a result of the generation of Big Data as enabled by networked devices, increased capabilities to analyze Big Data, and the prevalent use of networked systems - the principle of collection limitation is changing [10].
2. **Consent**: As a result of the use of data from a wide variety of sources and the re-use of data which is inherent in Big Data practices - notions of informed consent (initial and secondary) are changing [11].
3. **Data Minimization**: As a result of Big Data practices inherently utilizing all data possible - the principle of data minimization is changing/obsolete [12].
4. **Notice**: As a result of Big Data practices relying on vast amounts of data from numerous sources and the re-use of that data - the principle of notice is changing [13].
5. **Purpose Limitation**: As a result of Big Data practices re-using data for multiple purposes - the principle of purpose limitation is changing/obsolete [14].
6. **Necessity**: As a result of Big Data practices re-using data, the new use or re-analysis of data may not be pertinent to the purpose that was initially specified- thus the principle of necessity is changing [15].
7. **Access and Correction**: As a result of Big Data being generated (and sometimes published) at scale and in real time - the principle of user access and correction is changing [16].
8. **Opt In and Opt Out Choices**: Particularly in the context of smart cities and IoT which collect data on a real time basis, often without the knowledge of the individual, and for the provision of a service - it may not be easy or possible for individuals to opt in or out of the collection of their data [17].
9. **PI**: As a result of Big Data analytics using and analyzing a wide variety of data, new or unexpected forms of personal data may be generated - thus challenging and evolving beyond traditional or specified definitions of personal information [18].
10. **Data Controller**: In the context of IoT, given the multitude of actors that can collect, use and process data generated by networked devices, the traditional understanding of what and who is a data controller is changing [19].

Possible Technical and Policy Solutions

In a Report titled "*Internet of Things: Privacy & Security in a Connected World*" by the Federal Trade Commission in the United States it was noted that though IoT changes the application and understanding of certain privacy principles, it does not necessarily make them obsolete [20]. Indeed many possible solutions that have been suggested to address the challenges posed by IoT and Big Data are technical interventions at the device level rather than fundamental policy changes. For example it has been proposed that IoT devices can be programmed to:

- Automatically delete data after a specified period of time [21] (addressing concerns of data retention)
- Ensure that personal data is not fed into centralized databases on an automatic basis [22] (addressing concerns of transfer and sharing without consent, function creep, and data breach)
- Offer consumers combined choices for consent rather than requiring a one time blanket consent at the time of initiating a service or taking fresh consent for every change that takes place while a consumer is using a service [23]. (addressing concerns of informed and meaningful consent)
- Categorize and tag data with accepted uses and programme automated processes to flag when data is misused [24]. (addressing concerns of misuse of data)
- Apply 'sticky policies' - policies that are attached to data and define appropriate uses of the data as it 'changes hands' [25] (addressing concerns of user control of data)
- Allow for features to only be turned on with consent from the user [26] (addressing concerns of informed consent and collection without the consent or knowledge of the user)
- Automatically convert raw personal data to aggregated data [27] (addressing concerns of misuse of personal data and function creep)
- Offer users the option to delete or turn off sensors [28] (addressing concerns of user choice, control, and consent)

Such solutions place the designers and manufacturers of IoT devices in a critical role. Yet some, such as Kate Crawford and Jason Shultz are not entirely optimistic about the possibility of effective technological solutions - noting in the context of automated decision making that it is difficult to build in privacy protections as it is unclear when an algorithm will predict personal information about an individual [29].

Experts have also suggested that more emphasis should be placed on the principles and practices of:

- Transparency,
- Access and correction,
- Use/misuse
- Breach notification
- Remedy
- Ability to withdraw consent

Others have recommended that certain privacy principles need to be adapted to the Big Data/IoT context. For example, the Article 29 Working Party has clarified that in the context of IoT, consent mechanisms need to include the types of data collected, the frequency of data collection, as well as conditions for data collection [30]. While the Federal Trade Commission has warned that adopting a pure "use" based model has its limitations as it requires a clear (and potentially changing) definition of what use is acceptable and what use is not acceptable, and it does not address concerns around the collection of sensitive personal information [31]. In addition to the above, the European Commission has stressed that the right of deletion, the right to be forgotten, and data portability also need to be foundations of IoT systems and devices [32].

Possible Regulatory Frameworks

To the question - are current regulatory frameworks adequate and is additional legislation needed, the FTC has recommended that though a specific IoT legislation may not be

necessary, a horizontal privacy legislation would be useful as sectoral legislation does not always account for the use, sharing, and reuse of data across sectors. The FTC also highlighted the usefulness of privacy impact assessments and self regulatory steps to ensure privacy [33]. The European Commission on the other hand has concluded that to ensure enforcement of any standard or protocol - hard legal instruments are necessary [34]. As mentioned earlier, Kate Crawford and Jason Shultz have argued that privacy regulation needs to move away from principles on collection, specific use, disclosure, notice etc. and focus on elements of due process around the use of Big Data - as they say "procedural data due process". Such due process should be based on values instead of defined procedures and should include at the minimum notice, hearing before an independent arbitrator, and the right to review. Crawford and Shultz more broadly note that there are conceptual differences between privacy law and big data that pose as serious challenges i.e privacy law is based on causality while big data is a tool of correlation. This difference raises questions about how effective regulation that identifies certain types of information and then seeks to control the use, collection, and disclosure of such information will be in the context of Big Data – something that is varied and dynamic. According to Crawford and Shultz many regulatory frameworks will struggle with this difference – including the FTC's Fair Information Privacy Principles and the EU regulation including the EU's right to be forgotten [35]. The European Data Protection Supervisor on the other hand looks at Big Data as spanning the policy areas of data protection, competition, and consumer protection – particularly in the context of 'free' services. The Supervisor argues that these three areas need to come together to develop ways in which the challenges of Big Data can be addressed. For example, remedy could take the form of data portability – ensuring users the ability to move their data to other service providers empowering individuals and promoting competitive market structures or adopting a 'compare and forget' approach to data retention of customer data. The Supervisor also stresses the need to promote and treat privacy as a competitive advantage, thus placing importance on consumer choice, consent, and transparency [36]. The European Data Protection reform has been under discussion and it is predicted to be enacted by the end of 2015. The reform will apply across European States and all companies operating in Europe. The reform proposes heavier penalties for data breaches, seeks to provide users with more control of their data [37]. Additionally, Europe is considering bringing digital platforms under the Network and Information Security Directive – thus treating companies like Google and Facebook as well as cloud providers and service providers as a critical sector. Such a move would require companies to adopt stronger security practices and report breaches to authorities [38].

Conclusion

A review of the different opinions and reactions from experts and policy makers demonstrates the ways in which Big Data and IoT are changing traditional forms of protection that governments and societies have developed to protect personal data as it increases in value and importance. While some policy makers believe that big data needs strong legislative regulation and others believe that softer forms of regulation such as self or co-regulation are more appropriate, what is clear is that Big Data is either creating a regulatory dilemma– with policy makers searching for ways to control the unpredictable nature of big data through policy and technology through the merging of policy areas, the honing of existing policy mechanisms, or the broadening of existing policy mechanisms - while others are ignoring the change that Big Data brings with it and are forging ahead with its use.

Answering the 'how do we regulate Big Data" question requires **re-conceptualization of data ownership and realities**. Governments need to first recognize the criticality of their data and the data of their citizens/residents, as well as the contribution to a country's economy and security that this data plays. With the technologies available now, and in the pipeline, data can be used or misused in ways that will have vast repercussions for individuals, society, and a nation. All data, but especially data directly or indirectly related to citizens and residents of a country, needs to be looked upon as owned by the citizens and the nation. In this way,

data should be seen as a part of **critical national infrastructure of a nation**, and accorded the security, protections, and legal backing thereof to **prevent the misuse of the resource by the private or public sectors, local or foreign governments**. This could allow for local data warehousing and bring physical and access security of data warehouses on par with other critical national infrastructure. Recognizing data as a critical resource answers in part the concern that experts have raised – that Big Data practices make it impossible for data to be categorized as personal and thus afforded specified forms of protection due to the unpredictable nature of big data. Instead – all data is now recognized as critical.

In addition to being able to generate personal data from anonymized or non-identifiable data, big data also challenges traditional divisions of public vs. private data. Indeed Big Data analytics can take many public data points and derive a private conclusion. The use of Big Data analytics on public data also raises questions of consent. For example, though a license plate is public information – should a company be allowed to harvest license plate numbers, combine this with location, and sell this information to different interested actors? This is currently happening in the United States [39]. Lastly, Big Data raises questions of ownership. A solution to the uncertainty of public vs. private data and associated consent and ownership could be the creation a **National Data Archive** with such data. The archive could function with representation from the government, public and private companies, and civil society on the board. In such a framework, for example, companies like Airtel would provide mobile services, but the CDRs and customer data collected by the company would belong to the National Data Archive and be available to Airtel and all other companies within a certain scope for use. This 'open data' approach could enable innovation through the use of data but within the ambit of national security and concerns of citizens – a framework that could instill trust in consumers and citizens. Only when backed with strong security requirements, enforcement mechanisms and a proactive, responsive and responsible framework can governments begin to think about ways in which Big Data can be harnessed.

ENDNOTES

[1] BCS - The Chartered Institute for IT. (2013). The Societal Impact of the Internet of Things. Retrieved May 17, 2015, from http://www.bcs.org/upload/pdf/societal-impact-report-feb13.pdf

[2] Sicular, S. (2013, March 27). Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused with Three "V"s. Retrieved May 20, 2015, from http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/

[3] Executive Office of the President. "Big Data: Seizing Opportunities, Preserving Values". May 2014. Available at: https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf. Accessed: July 2nd 2015.

[4] Moses, B., Lyria, & Chan, J. (2014). Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools (SSRN Scholarly Paper No. ID 2513564). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2513564

[5] Danah Boyd, Kate Crawford. CRITICAL QUESTIONS FOR BIG DATA. Information, Communication & Society Vol. 15, Iss. 5, 2012. Available at: http://www.tandfonline.com/doi/full/10.1080/1369118X.2012.678878. Accessed: July 2nd2015.

[6] The Chartered Institute for IT, Oxford Internet Institute, University of Oxford. "The Societal Impact of the Internet of Things" February 2013. Available at: http://www.bcs.org/upload/pdf/societal-impact-report-feb13.pdf. Accessed: July 2nd 2015.

[7] ARTICLE 29 Data Protection Working Party. (2014). Opinion 8/2014 on the on Recent Developments on the Internet of Things. European Commission. Retrieved May 20, 2015, from http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf

[8] Crawford, K., & Schultz, J. (2013). Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms (SSRN Scholarly Paper No. ID 2325784). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2325784

[9] Barocas, S., & Selbst, A. D. (2015). Big Data's Disparate Impact (SSRN Scholarly Paper No. ID 2477899). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2477899

[10] Barocas, S., & Selbst, A. D. (2015). Big Data's Disparate Impact (SSRN Scholarly Paper No. ID 2477899). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2477899

[11] Article 29 Data Protection Working Party. “Opinion 8/2014 on the on Recent Developments on the Internet of Things”. September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[12] Tene, O., & Polonetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics. Northwestern Journal of Technology and Intellectual Property, 11(5), 239.

[13] Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 11 Nw. J. Tech. & Intell. Prop. 239 (2013).

[14] Article 29 Data Protection Working Party. “Opinion 8/2014 on the on Recent Developments on the Internet of Things”. September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[15] Information Commissioner’s Office. (2014). Big Data and Data Protection. Infomation Commissioner’s Office. Retrieved May 20, 2015, from https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf

[16] Article 29 Data Protection Working Party. “Opinion 8/2014 on the on Recent Developments on the Internet of Things”. September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[17] The Chartered Institute for IT and Oxford Internet Institute, University of Oxford. “The Societal Impact of the Internet of Things”. February 14th 2013. Available at: http://www.bcs.org/upload/pdf/societal-impact-report-feb13.pdf. Accessed: July 2nd 2015.

[18] Kate Crawford and Jason Shultz, “Big Data and Due Process: Towards a Framework to Redress Predictive Privacy Harms”. Boston College Law Review, Volume 55, Issue 1, Article 4. January 1st 2014. Available at: http://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3351&context=bclr. Accessed: July 2nd 2015.

[19] Article 29 Data Protection Working Party “Opinion 8/2014 on the on Recent Developments on the Internet of Things” September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[20] Federal Trade Commission. (2015). Internet of Things: Privacy & Security in a Connected World. Federal Trade Commision. Retrieved May 20, 2015, from https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

[21] Federal Trade Commission. (2015). Internet of Things: Privacy & Security in a Connected World. Federal Trade Commision. Retrieved May 20, 2015, from https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

[22] Federal Trade Commission. (2015). Internet of Things: Privacy & Security in a Connected World. Federal Trade Commision. Retrieved May 20, 2015, from https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

[23] Federal Trade Commission. (2015). Internet of Things: Privacy & Security in a Connected World. Federal Trade Commision. Retrieved May 20, 2015, from https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

[24] Federal Trade Commission. (2015). Internet of Things: Privacy & Security in a Connected World. Federal Trade Commision. Retrieved May 20, 2015, from https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

[25] Article 29 Data Protection Working Party “Opinion 8/2014 on the on Recent Developments on the Internet of Things” September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[26] Article 29 Data Protection Working Party “Opinion 8/2014 on the on Recent Developments on the Internet of Things” September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[27] Article 29 Data Protection Working Party “Opinion 8/2014 on the on Recent Developments on the Internet of Things” September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[28] Article 29 Data Protection Working Party “Opinion 8/2014 on the on Recent Developments on the Internet of Things” September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[29] Kate Crawford and Jason Shultz, “Big Data and Due Process: Towards a Framework to Redress Predictive Privacy Harms”. Boston College Law Review, Volume 55, Issue 1, Article 4. January 1st 2014. Available at: http://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3351&context=bclr. Accessed: July 2nd 2015.

[30] Article 29 Data Protection Working Party “Opinion 8/2014 on the on Recent Developments on the Internet of Things” September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[31] Federal Trade Commission. (2015). Internet of Things: Privacy & Security in a Connected World. Federal Trade Commission. Retrieved May 20, 2015, from https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

[32] Article 29 Data Protection Working Party “Opinion 8/2014 on the on Recent Developments on the Internet of Things” September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[33] Federal Trade Commission. (2015). Internet of Things: Privacy & Security in a Connected World. Federal Trade Commission. Retrieved May 20, 2015, from https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

[34] Article 29 Data Protection Working Party “Opinion 8/2014 on the on Recent Developments on the Internet of Things” September 16th 2014. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed: July 2nd 2015.

[35] Kate Crawford and Jason Shultz, “Big Data and Due Process: Towards a Framework to Redress Predictive Privacy Harms”. Boston College Law Review, Volume 55, Issue 1, Article 4. January 1st 2014. Available at: http://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3351&context=bclr. Accessed: July 2nd 2015.

[36] European Data Protection Supervisor. Preliminary Opinion of the European Data Protection Supervisor, Privacy and competitiveness in the age of big data: the interplay between data protection, competition law and consumer protection in the Digital Economy. March 2014. Available at: https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2014/14-03-26_competitition_law_big_data_EN.pdf

[37] SC Magazine. Harmonised EU data protection and fines by the end of the year. June 25th 2015. Available at: http://www.scmagazineuk.com/harmonised-eu-data-protection-and-fines-by-the-end-of-the-year/article/422740/. Accessed: August 8th 2015.

[38] Tom Jowitt, “Digital Platforms to be Included in EU Cybersecurity Law”. TechWeek Europe. August 7th 2015. Available at: http://www.techweekeurope.co.uk/e-regulation/digital-platforms-eu-cybersecuity-law-174415

[39] Adam Tanner. Data Brokers are now Selling Your Car’s Location for $10 Online. July 10th 2013. Available at: http://www.forbes.com/sites/adamtanner/2013/07/10/data-broker-offers-new-service-showing-where-they-have-spotted-your-car/

Big Data and the Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules 2011

ELONNAI HICKOK

Experts and regulators across the globe are examining the impact of Big Data practices on traditional data protection standards and principles. This will be a useful and pertinent exercise for India to undertake as the government and the private and public sectors begin to incorporate and rely on the use of Big Data in decision making processes and organizational operations.

Below is an initial evaluation of how Big Data could impact India's current data protection standards.

India currently does not have comprehensive privacy legislation - but the Reasonable Security Practices and Procedures and Sensitive Personal Data or Information Rules 2011 formed under section 43A of the Information Technology Act 2000 [1] define a data protection framework for the processing of digital data by Body Corporate. Big Data practices will impact a number of the provisions found in the Rules:

Scope of Rules

Currently the Rules apply to Body Corporate and digital data. As per the IT Act, Body Corporate is defined as "*Any company and includes a firm, sole proprietorship or other association of individuals engaged in commercial or professional activities.*"

The present scope of the Rules excludes from its purview a number of actors that do or could have access to Big Data or use Big Data practices. The Rules would not apply to government bodies or individuals collecting and using Big Data. Yet, with technologies such as IoT and the rise of Smart Cities across India – a range of government, public, and private organizations and actors could have access to Big Data.

Definition of Personal and Sensitive Personal Data

Rule 2(i) defines personal information as "*information that relates to a natural person which either directly or indirectly, in combination with other information available or likely to be available with a body corporate, is capable of identifying such person.*"

Rule 3 defines sensitive personal information as:

- Password,
- Financial information,
- Physical/physiological/mental health condition,
- Sexual orientation,
- Medical records and history,
- Biometric information

The present definition of personal data hinges on the factor of identification (data that is capable of identifying a person). Yet this definition does not encompass information that is associated to an already identified individual - such as habits, location, or activity.

The definition of personal data also addresses only the identification of 'such person' and does not address data that is related to a particular person but that also reveals identifying information about another person - either directly - or when combined with other data points.

By listing specific categories of sensitive personal information, the Rules do not account for additional types of sensitive personal information that might be generated or correlated through the use of Big Data analytics.

Importantly, the definitions of sensitive personal information or personal information do not address how personal or sensitive personal information - when anonymized or aggregated – should be treated.

Consent

Rule 5(1) requires that Body Corporate must, prior to collection, obtain consent in writing through letter or fax or email from the provider of sensitive personal data regarding the use of that data.

In a context where services are delivered with little or no human interaction, data is collected through sensors, data is collected on a real time and regular basis, and data is used and re-used for multiple and differing purposes - it is not practical, and often not possible, for consent to be obtained through writing, letter, fax, or email for each instance of data collection and for each use.

Notice of Collection

Rule 5(3) requires Body Corporate to provide the individual with a notice during collection of information that details the fact that information is being collected, the purpose for which the information is being collected, the intended recipients of the information, the name and address of the agency that is collecting the information and the agency that will retain the information. Furthermore body corporate should not retain information for longer than is required to meet lawful purposes.

Though this provision acts as an important element of transparency, in the context of Big Data, communicating the purpose for which data is collected, the intended recipients of the information, the name and address of the agency that is collecting the information and the agency that will retain the information could prove to be difficult to communicate as they are likely to encompass numerous agencies and change depending upon the analysis being done.

Access and Correction

Rule 5(6) provides individuals with the ability to access sensitive personal information held by the body corporate and correct any inaccurate information.

This provision would be difficult to implement effectively in the context of Big Data as vast amounts of data are being generated and collected on an ongoing and real time basis and often without the knowledge of the individual.

Purpose Limitation

Rule 5(5) requires that body corporate should use information only of the purpose which it has been collected.

In the context of Big Data this provision would overlook the re-use of data that is inherent in such practices.

Security

Rule 8 states that any Body Corporate or person on its behalf will be understood to have complied with reasonable security practices and procedures if they have implemented such practices and have in place codes that address managerial, technical, operational and physical security control measures. These codes could follow the IS/ISO/IEC 27001 standard or another government approved and audited standard.

This provision importantly requires that data controllers collecting and processing data have in place strong security practices. In the context of Big Data – the security of devices that might be generating or collecting data and algorithms processing and analysing data is critical. Once generated, it might be challenging to ensure the data is being transferred to or being analysed by organisations that comply with such security practices as listed.

Data Breach

Rule 8 requires that if a data breach occurs, Body Corporate would have to be able to demonstrate that they have implemented their documented information security codes.

Though this provision holds a company accountable for the implementation of security practices, it does not address how a company should be held accountable for a large scale data breach as in the context of Big Data the scope and impact of a data breach is on a much larger scale.

Opt In and Out and Ability to Withdraw Consent

Rule 5(7) requires Body Corporate or any person on its behalf, prior to the collection of information - including sensitive personal information - must give the individual the option of not providing information and must give the individual the option of withdrawing consent. Such withdrawal must be sent in writing to the body corporate.

The feasibility of such a provision in the context of Big Data is unclear, especially in light of the fact that Big Data practices draw upon large amounts of data, generated often in real time, and from a variety of sources.

Disclosure of Information

Rule 6 maintains that disclosure of sensitive personal data can only take place with permission from the provider of such information or as agreed to through a lawful contract.

This provision addresses disclosure and does not take into account the “sharing” of information that is enabled through networked devices, as well as the increasing practice of companies to share anonymized or aggregated data.

Privacy Policy

Rule 4 requires that body corporate have in place a privacy policy on their website that provides clear and accessible statements of its practices and policies, type of personal or sensitive personal information that is being collected, purpose of the collection, usage of the information, disclosure of the information, and the reasonable security practices and procedures that have been put in place to secure the information.

In the context of Big Data where data from a variety of sources is being collected, used, and re-used it is important for policies to ‘follow data’ and appear in a contextualized manner. The current requirement of having Body Corporate post a single overarching privacy policy on its website could prove to be inadequate.

Remedy

Section 43A of the Act holds that if a body corporate is negligent in implementing and maintain reasonable security practices and procedures which results in wrongful loss or wrongful gain to any person, the body corporate can be held liable to pay compensation to the affected person.

This provision will provide limited remedy for an affected individual in the context of Big Data. Though important to help prevent data breaches resulting from negligent data practices, implementation of reasonable security practices and procedures cannot be the only hinging point for determining liability of a Body Corporate for violations and many of the harms possible through Big Data are not in the form of wrongful loss or wrongful gain to another person. Indeed many harms possible through Big Data are non-economic in nature – including physical invasion of privacy, and discriminatory practices that can arise from decisions based on Big Data analytics. Nor does the provision address the potential for future damage that can result from a 'Big Data data breach'.

The safeguards noted in the above section are not the only legal provisions that speak to privacy in India. There are over fifty sectoral legislation that have provisions addressing privacy - for example provisions addressing confidentiality of health and banking information. The government of India is also in the process of drafting a privacy legislation. In 2012 the Report of the Group of Experts on Privacy provided recommendations for a privacy framework in India. The Report envisioned a framework of co-regulation - with sector level self regulatory organization developing privacy codes (that are not lower than the defined national privacy principles) and that are enforced by a privacy commissioner [2]. Perhaps this method would be optimal for the regulation of Big Data- allowing for the needed flexibility and specificity in standards and device development. Though the Report notes that individuals can seek remedy from the court and the Privacy Commissioner can issue fines for a violation, the development of privacy legislation in India has yet to clearly integrate the importance of due process and remedy. With the onset of Big Data - this will become more important than ever.

Conclusion

The use and generation of Big Data in India is growing. Plans such as free wifi zones in cities [3], city wide CCTV networks with facial recognition capabilities [4], and the implementation of an identity/authentication platform for public and private services [5], are indicators towards a move of data generation that is networked and centralized, and where the line between public and private is blurred through the vast amount of data that is collected.

In such developments and innovations what is privacy and what role does privacy play? Is it the archaic inhibitor - limiting the sharing and use of data for new and innovative purposes? Will it be defined purely by legislative norms or through device/platform design as well? Is it a notion that makes consumers think twice about using a product or service or is it a practice that enables consumer and citizen uptake and trust and allows for the growth and adoption of these services?

How privacy will be regulated and how it will be perceived is still evolving across jurisdictions, technologies, and cultures - but it is clear that privacy is not being and cannot be overlooked. Governments across the world are reforming and considering current and future privacy regulation targeted towards life in a quantified society. As the Indian government begins to roll out initiatives that create a "Digital India" indeed a "quantified India", taking privacy into consideration could facilitate the uptake, expansion, and success of these practices and services. As the Indian government pursues the opportunities possible through Big Data it will be useful to review existing privacy protections and deliberate on if, and in what form, future protections for privacy and other rights will be needed.

ENDNOTES

[1] Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information Rules 2011). Available at: http://deity.gov.in/sites/upload_files/dit/files/GSR313E_10511(1).pdf

[2] Group of Experts on Privacy. (2012). *Report of the Group of Experts on Privacy*. New Delhi: Planning Commission, Government of India. Retrieved May 20, 2015, from http://planningcommission.nic.in/reports/genrep/rep_privacy.pdf

[3] NDTV. "Free Public Wi-Fi Facility in Delhi to Have Daily Data Limit. NDTV, May 25th 2015, Available at: http://gadgets.ndtv.com/internet/news/free-public-wi-fi-facility-in-delhi-to-have-daily-data-limit-695857. Accessed: July 2nd 2015.

[4] FindBiometrics Global Identity Management. "Surat Police Get NEC Facial Recognition CCTV System". July 21st2015. Available at: http://findbiometrics.com/surat-police-nec-facial-recognition-27214/

[5] UIDAI Official Website. Available at: https://uidai.gov.in/

Too Clever by Half: Strengthening India's Smart Cities Plan with Human Rights Protection

VANYA RAKESH

As Indian cities reposition themselves to play a significant role in development due to urban transformation, the government has envisioned building 100 smart cities across the country. Due to the lack of a precise definition as to what exactly constitutes a smart city, the mutual consensus that has evolved is that modern technology will be harnessed, which will lead to smart outcomes.

Here, Big Data and analytics will play a predominant role by the way of cloud, mobile technology and other social technologies that gather data for the purpose of ascertaining and accordingly addressing concerns of people.

Role of Big Data

Leveraging city data and using geographical information systems (GIS) to collect valuable information about stakeholders are some techniques that are commonly used in smart cities to execute emergency systems, creating dynamic parking areas, naming streets, and develop monitoring. Other sources which would harness such data would be from fire alarms, in disaster management situations and energy saving mechanisms, which would sense, communicate, analyze and combine information across platforms to generate data to facilitate decision making and manage services.

According to the Department of Electronics and Information Technology, the government's plan to develop smart cities in the country could lead to a massive expansion of an IoT (Internet of Things) ecosystem within the country. The revised draft IoT policy aims at developing IoT products in this domain by using Big Data for government decision-making processes. For example, in India a key opportunity that has been identified is with regard to traffic management and congestion. Here, collecting data during peak hours, processing information in real time and using GPS history from mobile phones can give insight into the routes taken and modes of transportation preferred by commuters to deal with traffic woes. The Bengaluru Transport Information System (BTIS) was an early adopter of big data technology which resorted to aggregating data streams from multiple sources to enable planning of travel routes by avoiding traffic congestions, car-pooling, etc.

Challenges

The idea of a data-driven urban city has drawn criticism as the initiative tends to homogenize Indian culture and change the fabric of cities by treating them alike in terms of their political economy, culture, and governance.

Despite basing the idea of a smart city on the assumption that technology-based solutions and techniques would be a viable solution for city problems in India, it is pertinent to note that the collection of personal real-time data may blur the line between personal data with the large data collected from multiple sources, leaving questions around privacy considerations, use and reuse of such data, especially by companies and businesses involved in providing services in legally and morally grey areas.

Privacy concerns cloud the dependence on big data for functioning of smart cities as it may lead to erosion of privacy in different forms, for example if it is used to carry out surveillance, identification and disclosures without consent, discriminatory inferences, etc.

Apart from right to privacy, a number of rights of an individual like the right to access and security rights would be at risk as it may enable practices of algorithmic social sorting (whether people get a loan, a tenancy, a job, etc.), and anticipatory governance using predictive profiling (wherein data precedes how a person is policed and governed). Dataveillance raises concerns around access and use of data due to increase in digital footprints (data they themselves leave behind) and data shadows (information about them generated by others). Also, the challenges and the realities of getting access to correct and standardized data, and proper communication seem to be a hurdle which still needs to be overcome.

The huge, yet untapped, amount of data available in India requires proper categorization and this makes a robust and reliable data management system prerequisite for realization of the country's smart city vision. Cooperation between agencies in Indian cities and a holistic technology-based approach like ICT and GT (geospatial technologies) to resolve issues pertaining to wide use of technology is the need of the hour. The skills to manage, analyze and develop insights for effective policy decisions are still being developed, particularly in the public sector. Recognizing this, Nasscom in India has announced setting up a Centre of Excellence (CoE) to create quality workforce.

Though it is apparent that data will play a considerable role in smart city mission, the peril is lack of planning in terms of policies to govern the big data mechanics and use of data. This calls for development of suitable standards and policies to guide technology providers & administrators to manage and interpret data in a secured environment.

Legal Hurdles

The Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules 2011 deals with accountability regarding data security and protection as it applies to 'body corporates' and digital data. It defines a 'body corporate' as "any company and includes a firm, sole proprietorship or other association of individuals engaged in commercial or professional activities" under the IT Act. Therefore, it can be ascertained that government bodies or individuals collecting and using Big Data for the smart cities in India would be excluded from the scope of these Rules. This highlights the lack of a suitable regulatory framework to take into account potential privacy challenges, which currently seem to be underestimated by our planners and administrators.

Regarding access to open data, though the National Data Sharing and Accessibility Policy 2012 recognizes sensitive data, the term has not been clearly defined under it. However, the Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules, 2011 clearly define sensitive personal data or information. Therefore, the open data framework must refer to or adopt a clear definition drawing from section 43A Rules to bring clarity in this regard.

Way Forward

As India moves toward a digital transformation, highlighted by flagship programmes like Smart Cities Mission, Digital India and the UID project, data regulation and recognition of use of data will change the nature of the relationship between the state and the individual. However, this seems to have been overlooked. Policies that regulate the digital environment of the country will intertwine with urban policies due to the smart cities mission. Use of ICTs in the form of IoT and Big Data entails access to open data, bringing another policy area in its ambit which needs consideration. Identification/development of open standards for IoT particularly for interoperability between cross sector data must be looked at.

To address privacy concerns due to the use of big data techniques, nuanced data legislation is required. For a conducive big data and technologically equipped environment, the governments must increase efforts to create awareness about the risks involved and provide assurance about the responsible use of data.

Additionally, a lack of skilled and educated manpower to deal with such data effectively must also be duly considered.

The concept note produced by the government reflects how it visualizes smart cities to be a product of marrying the physical form of cities and its infrastructure to a wider discourse on the use of technology and big data in city governance. This makes the role of big data quite indispensable, making it synonymous with the very notion of a smart city. However, the important issue is to understand that data analytics is only a part of the idea. What is additionally required is effective governance mechanism and political will. Collaboration and co-operation is the glue that will make this idea work. It is important to merge urban development policies with principles of democracy. The data involved in planning for urbanized and networked cities are currently flawed and politically-inflected. Therefore, collective efforts must go into minimizing pernicious effects of the same to ensure the basic human rights are not violated in the race to make cities “smart”.

Predictive Policing: What is it, How it Works, and its Legal Implications

ROHAN GEORGE

Introduction

For the longest time, humans have been obsessed with prediction. Perhaps the most well-known oracle in history, Pythia, the infallible Oracle of Delphi was said to predict future events in hysterical outbursts on the seventh day of the month, inspired by the god Apollo himself. This fascination with informing ourselves about future events has hardly subsided in us humans. What has changed however is the methods we employ to do so. The development of Big data technologies for one, has seen radical applications into many parts of life as we know it, including enhancing our ability to make accurate predictions about the future.

One notable application of Big data into prediction caters to another basic need since the dawn of human civilisation, the need to protect our communities and cities. The word 'police' itself originates from the Greek word 'polis', which means city. The melding of these two concepts prediction and policing has come together in the practice of Predictive policing, which is the application of computer modelling to historical crime data and metadata to predict future criminal activity [1]. In the subsequent sections, I will attempt an introduction of predictive policing and explain some of the main methods within the domain of predictive policing. Because of the disruptive nature of these technologies, it will also be prudent to expand on the implications predictive technologies have for justice, privacy protections and protections against discrimination among others.

In introducing the concept of predictive policing, my first step is to give a short explanation about current predictive analytics techniques, because these techniques are the ones which are applied into a law enforcement context as predictive policing.

What is Predictive Analysis

Facilitated by the availability of big data, predictive analytics uses algorithms to recognise data patterns and predict future outcomes [2]. Predictive analytics encompasses data mining, predictive modeling, machine learning, and forecasting [3]. Predictive analytics also relies heavily on machine learning and artificial intelligence approaches [4]. The aim of such analysis is to identify relationships among variables that may not be immediately apparent using hypothesis-driven methods [5]. In the mainstream media, one of the most infamous stories about the use of predictive analysis comes from USA, regarding a department store Target and their data analytics practices [6]. Target mined data from purchasing patterns of people who signed onto their baby registry. From this they were able to predict approximately when customers may be due and target advertisements accordingly. In the noted story, they were so successful that they predicted pregnancy before the pregnant girl's father knew she was pregnant [7].

Examples of predictive analytics

- Predicting the success of a movie based on its online ratings [8]
- Many universities, sometimes in partnership with other firms use predictive analytics to provide course recommendations to students, track student performance, personalize curriculum to individual students and foster networking between students [9].
- Predictive Analysis of Corporate Bond Indices Returns [10]

Relationship Between Predictive Analytics and Predictive Policing

The same techniques used in many of the predictive methods mentioned above find application into some predictive policing methods. However two important points need to be raised:

First, predictive analytics is actually a subset of predictive policing. This is because while the steps in creating a predictive model, of defining a target variable, exposing your model to training data, selecting appropriate features and finally running predictive analysis [11] maybe the same in a policing context, there are other methods which may be used to predict crime, but which do not rely on data mining. These techniques may instead use other methods, such as some of those detailed below along with data about historical crime to generate predictions.

In her article "Policing by Numbers: Big Data and the Fourth Amendment" [12], Joh categorises 3 main applications of Big data into policing. These are Predictive Policing, Domain Awareness systems and Genetic Data Banks. Genetic data banks refer to maintaining large databases of DNA that was collected as part of the justice system. Issues arise when the DNA collected is repurposed in order to conduct familial searches, instead of being used for corroborating identity. Familial searches may have disproportionate impacts on minority races. Domain Awareness systems use various computer software and other digital surveillance tools such as Geographical Information Systems [13] or more illicit ones such as Black Rooms [14] to "help police create a software-enhanced picture of the present, using thousands of data points from multiple sources within a city" [15]. I believe Joh was very accurate in separating Predictive Policing from Domain Awareness systems, especially when it comes to analysing the implications of the various applications of Big data into policing.

In such an analysis of the implications of using predictive policing methods, the issues surrounding predictive technologies often get conflated with larger issues about the application of big data into law enforcement. That opens the debate up to questions about overly intrusive evidence gathering and mass surveillance systems, which though used along with predictive technology, are not themselves predictive in nature. In this article, I aim to concentrate on the specific implications that arise due to predictive methods.

One important point regarding the impact of predictive policing is how the insights that predictive policing methods offer are used. There is much support for the idea that predictive policing does not replace policing methods, but actually augments them. The RAND report specifically cites one myth about predictive policing as "the computer will do everything for you [16]". In reality police officers need to act on the recommendations provided by the technologies.

What is Predictive policing?

Predictive policing is the "application of analytical techniques-particularly quantitative techniques-to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions" [17]. It is important to note that the use of data and statistics to inform policing is not new. Indeed, even twenty years ago, before the deluge of big data we have today, law enforcement regimes such as the New York Police Department (NYPD) were already using crime data in a major way. In order to keep track of crime trends, NYPD used the software CompStat [18] to map "crime statistics along with other indicators of problems, such as the locations of crime victims and gun arrests" [19]. The senior officers used the information provided by CompStat to monitor trends of crimes on a daily basis and such monitoring became an instrumental way to track the performance of police agencies [20]. CompStat has since seen application in many other jurisdictions [21].

But what is new is the amount of data available for collection, as well as the ease with which organisations can analyse and draw insightful results from that data. Specifically, new technologies allow for far more rigorous interrogation of data and wide-ranging applications, including adding greater accuracy to the prediction of future incidence of crime.

Predictive Policing Methods

Some methods of predictive policing involve application of known standard statistical methods, while other methods involve modifying these standard techniques. Predictive techniques that forecast future criminal activities can be framed around six analytic categories. They all may overlap in the sense that multiple techniques are used to create actual predictive policing software and in fact it is similar theories of criminology which undergird many of these methods, but the categorisation in such a way helps clarify the concept of predictive policing. The basis for the categorisation below comes from a RAND Corporation report entitled 'Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations' [22], which is a comprehensive and detailed contribution to scholarship in this nascent area.

Hot spot analysis: Methods involving hot spot analysis attempt to "predict areas of increased crime risk based on historical crime data" [23]. The premise behind such methods lies in the adage that "crime tends to be lumpy" [24]. Hot Spot analysis seeks to map out these previous incidences of crime in order to inform potential future crime.

Regression methods: A regression aims to find relationships between independent variables (factors that may influence criminal activity) and certain variables that one aims to predict. Hence, this method would track more variables than just crime history.

Data mining techniques: Data mining attempts to recognise patterns in data and use it to make predictions about the future. One important variant in the various types of data mining methods used in policing are different types of algorithms that are used to mine data in different ways. These are dependent on the nature of the data the predictive model was trained on and will be used to interrogate in the future. Two broad categories of algorithms commonly used are clustering algorithms and classification algorithms:

· Clustering algorithms "form a class of data mining approaches that seek to group data into clusters with similar attributes" [25]. One example of clustering algorithms is spatial clustering algorithms, which use geospatial crime incident data to predict future hot spots for crime [26].

· Classification algorithms "seek to establish rules assigning a class or label to events" [27]. These algorithms use training data sets "to learn the patterns that determine the class of an observation" [28] The patterns identified by the algorithm will be applied to future data, and where applicable, the algorithm will recognise similar patterns in the data. This can be used to make predictions about future criminal activity for example.

Near-repeat methods: Near-repeat methods work off the assumption that future crimes will take place close to timing and location of current crimes. Hence, it could be postulated that areas of high crime will experience more crime in the near future [29]. This involves the use of a 'self-exciting' algorithm, very similar to algorithms modelling earthquake aftershocks [30]. The premise undergirding such methods is very similar to that of hot spot analysis.

Spatiotemporal analysis: Using "environmental and temporal features of the crime location" [31] as the basis for predicting future crime. By combining the spatiotemporal features of the crime area with crime incident data, police could use the resultant information to predict the location and time of future crimes. Examples of factors that may be considered include timing of crimes, weather, distance from highways, time from payday and many more.

Risk terrain analysis: Analyses other factors that are useful in predicting crimes. Examples of such factors include "the social, physical, and behavioural factors that make certain areas more likely to be affected by crime" [32]

Various methods listed above are used, often together, to predict the where and when a crime may take place or even potential victims. The unifying thread which relates these methods is their dependence on historical crime data.

Examples of Predictive Policing:

Most uses of predictive policing that have been studied and reviewed in scholarly work come from the USA, though I will detail one case study from Derbyshire, UK. Below is a collation of various methods that are a practical application of the methods raised above.

Hot Spot analysis in Sacramento: In February 2011, Sacramento Police Department began using hot spot analysis along with research on optimal patrol time to act as a sufficient deterrent to inform how they patrol high-risk areas. This policy was aimed at preventing serious crimes by patrolling these predicted hot spots. In places where there was such patrolling, serious crimes reduced by a quarter with no significant increases such crimes in surrounding areas [33].

Data Mining and Hot Spot Mapping in Derbyshire, UK: The Safer Derbyshire Partnership, a group of law enforcement agencies and municipal authorities sought to identify juvenile crime hotspots [34]. They used MapInfo software to combine "multiple discrete data sets to create detailed maps and visualisations of criminal activity, including temporal and spatial hotspots" [35]. This information informed law enforcement about how to optimally deploy their resources.

Regression models in Pittsburgh: Researchers used reports from Pittsburgh Bureau of Police about violent crimes and "leading indicator" [36] crimes, crimes that were relatively minor but which could be a sign of potential future violent offences. The researcher ran analysis of areas with violent crimes, which were used as the dependent variable in analysing whether violent crimes in certain areas could be predicted by the leading indicator data. From the 93 significant violent crime areas that were studied, 19 areas were successfully predicted by the leading indicator data. [37]

Risk terrain modelling analysis in Morris County, New Jersey: Police in Morris County, used risk terrain analysis to tackle violent crimes and burglaries. They considered five inputs in their model: "past burglaries, the address of individuals recently arrested for property crimes, proximity to major highways, the geographic concentration of young men and the location of apartment complexes and hotels" [38]. The Morris County law enforcement officials linked the significant reductions in violent and property crime to their use of risk terrain modelling [39].

Near-repeat & hot spot analysis used by Santa Cruz Police Department: Uses PredPol software that applies the Mohler's algorithm [40] to a database with five years' worth of crime data to assess the likelihood of future crime occurring in the geographic areas within the city. Before going on shift, officers receive information identifying 15 such areas with the highest probability of crime [41]. The initiative has been cited as being very successful at reducing burglaries, and was used in Los Angeles and Richmond, Virginia [42].

Data Mining and Spatiotemporal analysis to predict future criminal activities in Chicago: Officers in Chicago Police Department made visits to people their software predicted were likely to be involved in violent crimes [43], guided by an algorithm-generated "Heat List" [44]. Some of the inputs used in the predictions include some types of arrest records, gun ownership, social networks [45] (police analysis of social networking is also a rising trend in predictive policing [46]) and generally type of people you are acquainted with [47] among others, but the full list of the factors are not public. The list sends police officers (or sometimes mails letters) to peoples' homes to offer social services or deliver warnings about the consequences for offending. Based in part on the information provided by the algorithm, officers may provide people on the Heat List information about vocational training programs or warnings about how Federal Law provides harsher punishments for reoffending [48].

Predictive Policing in India

In this section, I map out some of the developments in the field of predictive policing within India. On the whole, predictive policing is still very new in India, with Jharkhand being the

only state that appears to already have concrete plans in place to introduce predictive policing.

Jharkhand Police

The Jharkhand police began developing their IT infrastructure such as a Geographic Information System (GIS) and Server room when they received funding for Rs. 18.5 crore from the Ministry of Home Affairs [49]. The Open Group on E-governance (OGE), founded as a collaboration between the Jharkhand Police and National Informatics Centre [50], is now a multi-disciplinary group which takes on different projects related to IT [51]. With regards to predictive policing, some members of OGE began development in 2013 of data mining software which will scan online records that are digitised. The emerging crime trends "can be a building block in the predictive policing project that the state police want to try." [52]

The Jharkhand Police was also reported in 2012 to be in the final stages of forming a partnership with IIM-Ranchi [53]. It was alleged the Jharkhand police aimed to tap into IIM's advanced business analytics skills [54], skills that can be very useful in a predictive policing context. Mr Pradhan suggested that "predictive policing was based on intelligence-based patrol and rapid response"[55] and that it could go a long way to dealing with the threat of Naxalism in Jharkhand [56].

However, in Jharkhand, the emphasis appears to be targeted at developing a massive Domain Awareness system, collecting data and creating new ways to present that data to officers on the ground, instead of architecting and using predictive policing software. For example, the Jharkhand police now have in place "a Naxal Information System, Crime Criminal Information System (to be integrated with the CCTNS) and a GIS that supplies customised maps that are vital to operations against Maoist groups" [57]. The Jharkhand police's "Crime Analytics Dashboard" [58] shows the incidence of crime according to type, location and presents it in an accessible portal, providing up-to-date information and undoubtedly raises the situational awareness of the officers. Arguably, the domain awareness systems that are taking shape in Jharkhand would pave the way for predictive policing methods to be applied in the future. These systems and hot spot maps seem to be the start of a new age of policing in Jharkhand.

Predictive Policing Research

One promising idea for predictive policing in India comes from the research conducted by Lavanya Gupta and others entitled "Predicting Crime Rates for Predictive Policing"[59], which was a submission for the Gandhian Young Technological Innovation Award. The research uses regression modelling to predict future crime rates. Drawing from First Information Reports (FIRs) of violent crimes (murder, rape, kidnapping etc.) from Chandigarh Police, the team attempted "to extrapolate annual crime rate trends developed through time series models. This approach also involves correlating past crime trends with factors that will influence the future scope of crime, in particular demographic and macro-economic variables" [60]. The researchers used early crime data as the training data for their model, which after some testing, eventually turned out to have an accuracy of around 88.2%. [61] On the face of it, ideas like this could be the starting point for the introduction of predictive policing into India.

The rest of India's law enforcement bodies do not appear to be lagging behind. In the 44th All India police science congress, held in Gandhinagar, Gujarat in March this year, one of the Themes for discussion was the "Role of Preventive Forensics and latest developments in Voice Identification, Tele-forensics and Cyber Forensics" [62].Mr A K Singh, (Additional Director General of Police, Administration) the chairman of the event also said in an interview that there was to be a round-table DGs (Director General of Police) held at the conference to discuss predictive policing [63]. Perhaps predictive policing in India may not be that far away from reality.

CCTNS and the building blocks of Predictive policing

The Ministry of Home Affairs conceived of a Crime and Criminals Tracking and Network System (CCTNS) as part of national e-Governance plans. According to the website of the National Crime Records Bureau (NCRB), CCTNS aims to develop "a nationwide networked infrastructure for evolution of IT-enabled state-of-the-art tracking system around 'investigation of crime and detection of criminals' in real time" [64]

The plans for predictive policing seem in the works, but first steps that are needed in India across police forces involve digitizing data collection by the police, as well as connecting law enforcement agencies. The NCRB's website described the current possibility of exchange of information between neighbouring police stations, districts or states as being "next to impossible" [65]. The aim of CCTNS is precisely to address this gap and integrate and connect the segregated law enforcement arms of the state in India, which would be a foundational step in any initiatives to apply predictive methods.

What are the Implications of Using Predictive Policing? Lessons from USA

Despite the moves by law enforcement agencies to adopt predictive policing, one reality is that the implications of predictive policing methods are far from clear. This section will examine these implications on the carriage of justice and its use in law, as well as how it impacts privacy concerns for the individual. It frames the existing debates surrounding these issues with predictive policing, and aims to apply these principles into an Indian context.

Justice, Privacy & IV Amendment

Two key concerns about how predictive policing methods may be used by law enforcement relate to how insights from predictive policing methods are acted upon and how courts interpret them. In the USA, this issue may finds its place under the scope of IV Amendment jurisprudence. The IV amendment states that all citizens are "secure from unreasonable searches and seizures of property by the government" [66]. In this sense, the IV amendment forms the basis for search and surveillance law in the USA.

A central aspect of the IV Amendment jurisprudence is drawn from *United States v. Katz*. In *Katz*, the FBI attached a microphone to the outside of a public phone booth to record the conversations of Charles Katz, who was making phone calls related to illegal gambling. The court ruled that such actions constituted a search within the auspices of the 4th amendment. The ruling affirmed constitutional protection of all areas where someone has a "reasonable expectation of privacy" [67].

Later cases have provided useful tests for situations where government surveillance tactics may or may not be lawful, depending on whether it violates one's reasonable expectation of privacy. For example, in *United States v. Knotts*, the court held that "police use of an electronic beeper to follow a suspect surreptitiously did not constitute a Fourth Amendment search" [68]. In fact, some argue that that the Supreme Court's reasoning in such cases suggests " any 'scientific enhancement' of the senses used by the police to watch activity falls outside of the Fourth Amendment's protections if the activity takes place in public" [69]. This reasoning is based on the third party doctrine which holds that "if you voluntarily provide information to a third party, the IV Amendment does not preclude the government from accessing it without a warrant" [70]. The clearest exposition of this reasoning was in Smith v. Maryland, where the presiding judges noted that "this Court consistently has held that a person has no legitimate expectation of privacy in information he voluntarily turns over to third parties" [71].

However, the third party has seen some challenge in recent time. In *United States v. Jones*, it was ruled that the government's warrantless GPS tracking of his vehicle 24 hours a day for 28 days violated his Fourth Amendment rights [72]. Though the majority ruling was that warrantless GPS tracking constituted a search, it was in a concurring opinion written by Justice Sonya Sotomayor that such intrusive warrantless surveillance was said to infringe one's reasonable expectation of privacy. As Newell reflected on Sotomayor's opinion,

"Justice Sotomayor stated that the time had come for Fourth Amendment jurisprudence to discard the premise that legitimate expectations of privacy could only be found in situations of near or complete secrecy. Sotomayor argued that people should be able to maintain reasonable expectations of privacy in some information voluntarily disclosed to third parties" [73].

She said that the court's current reasoning on what constitutes reasonable expectations of privacy in information disclosed to third parties, such as email or phone records or even purchase histories, is "ill-suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks" [74].

Predictive Policing vs. Mass surveillance and Domain Awareness Systems

However, there is an important distinction to be drawn between these cases and evidence from predictive policing. This has to do with the difference in nature of the evidence collection. Arguably, from Jones and others, what we see is that use of mass surveillance and domain awareness systems, drawing from Joh's categorisation of domain awareness systems as being distinct from predictive policing mentioned above, could potentially encroach on one's reasonable expectation of privacy. However, I think that predictive policing, and the possible implications for justice associated with it, its predictive harms, are quite distinct from what has been heard by courts thus far.

The reason for distinct risks between predictive harms and privacy harms originating from information gathering is related to the nature of predictive policing technologies, and how they are used. It is highly unlikely that the evidence submitted by the State to indict an offender will be mainly predictive in nature. For example, would it be possible to convict an accused person solely on the premise that he was predicted to be highly likely to commit a crime, and that subsequently he did? The legal standard of proving guilt beyond a reasonable doubt [75] can hardly be met solely on predictive evidence for a multitude of reasons. Predictive policing methods could at most, be said to inform police about the risk of someone committing a crime or of crime happening at a certain location, as demonstrated above.

Predictive Policing and Criminal Procedure

It may therefore pay to analyse how predictive policing may be used across the various processes within the criminal justice system. In fact, in an analysis of the various stages of criminal procedure, from opening an investigation to gathering evidence, followed by arrest, trial, conviction and sentencing, we see that as the individual gets subject to more serious incursions or sanctions by the state, it takes a higher standard of certainty about wrongdoing and a higher burden of proof, in order to legitimize that particular action.

Hence, at more advanced stages of the criminal justice process such as seeking arrest warrants or trial, it is very unlikely that predictive policing on its own can have a tangible impact, because the nature of predictive evidence is probability based. It aims to calculate the risk of future crime occurring based on statistical analysis of past crime data [76]. While extremely useful, probabilities on their own will not come remotely close meet the legal standards of proving 'guilt beyond reasonable doubt'. It may be at the earlier stages of the criminal justice process that evidence predictive policing might see more widespread application, in terms of applying for search warrants and searching suspicious people while on patrol.

In fact, in the law enforcement context, prediction as a concept is not new to justice. Both courts and law enforcement officials already make predictions about future likelihood of crimes. In the case of issuing warrants, the IV amendment makes provisions that law enforcement officials show that the potential search is based "upon probable cause" [77] in order for a judge to grant a warrant. In *US v. Brinegar*, probable cause was defined as existing

"where the facts and circumstances within the officers' knowledge, and of which they have reasonably trustworthy information, are sufficient in themselves to warrant a belief by a man of reasonable caution that a crime is being committed" [78]. Again, this legal standard seems too high for predictive evidence meet.

However, the police also have an important role to play in preventing crimes by looking out for potential crimes while on patrol or while doing surveillance. When the police stop a civilian on the road to search him, reasonable suspicion must be established. This standard of reasonable suspicion was defined in most clearly in *Terry v. Ohio*, which required police to "be able to point to specific and articulable facts which, taken together with rational inferences from those facts, reasonably warrant that intrusion" [79]. Therefore, "reasonable suspicion that 'criminal activity may be afoot' is at base a prediction that the facts and circumstances warrant the reasonable prediction that a crime is occurring or will occur" [80]. Despite the assertion that "there are as of yet no reported cases on predictive policing in the Fourth Amendment context" [81], examining the impact of predictive policing on the doctrine of reasonable suspicion could be very instructive in understanding the implications for justice and privacy [82].

Predictive Policing and Reasonable Suspicion

Ferguson's insightful contribution to this area of scholarship involves the identification of existing areas where prediction already takes place in policing, and analogising them into a predictive policing context [83]. These three areas are: responding to tips, profiling, and high crime areas (hot spots).

Tips

Tips are pieces of information shared with the police by members of the public. Often tips, either anonymous or from known police informants, may predict future actions of certain people, and require the police to act on this information. The precedent for understanding the role of tips in probable cause comes from *Illinois v. Gates* [84]. It was held that "an informant's 'veracity,' 'reliability,' and 'basis of knowledge'-remain 'highly relevant in determining the value'" [85] of the said tip. Anonymous tips need to be detailed, timely and individualised enough [86] to justify reasonable suspicion [87]. And when the informant is known to be reliable, then his prior reliability may justify reasonable suspicion despite lacking a basis in knowledge [88].

Ferguson argues that whereas predictive policing cannot provide individualised tips, it is possible to consider reliable tips about certain areas as a parallel to predictive policing [89]. And since the courts had shown a preference for reliability even in the face of a weak basis in knowledge, it is possible to see the reasonable suspicion standard change in its application [90]. It also implies that IV protections may be different in places where crime is predicted to occur [91].

Profiling

Despite the negative connotations and controversial overtones at the mere sound of the word, profiling is already a method commonly used by law enforcement. For example, after a crime has been committed and general features of the suspect identified by witnesses, police often stop civilians who fit this description. Another example of profiling is common in combating drug trafficking [92], where agents keep track of travellers at airports to watch for suspicious behaviour. Based on their experience of common traits which distinguish drug traffickers from regular travellers (a profile), agents may search travellers if they fit the profile [93]. In the case of *United States v. Sokolow* [94], the courts "recognized that a drug courier profile is not an irrelevant or inappropriate consideration that, taken in the totality of circumstances, can be considered in a reasonable suspicion determination" [95]. Similar lines of thinking could be employed in observing people exchanging small amounts of money in an area known for high levels of drug activity, conceiving predictive actions as a form of profile [96].

It is valid to consider predictive policing as a form of profiling [97], but Ferguson argues that the predictive policing context means this 'new form' of profiling could change IV analysis. The premise behind such an argument lies in the fact that a prediction made by some algorithm about potential high risk of crime in a certain area, could be taken in conjunction observations of ordinarily innocuous events. Read in the totality of circumstances, these two threads may justify individual reasonable suspicion [98]. For example, a man looking into cars at a parking lot may not by itself justify reasonable suspicion, but taken together with a prediction of high risk of car theft at that locality, it may well justify reasonable suspicion. It is this impact of predictive policing, which influences the analysis of reasonable suspicion in a totality of circumstances that may represent new implications for courts looking at IV amendment protections.

Profiling, Predictive Policing and Discrimination

The above sections have already brought up the point that law enforcement agencies already utilize profiling methods in their operations. Also, as the sections on how predictive analytics works and on methods of predictive policing make clear, predictive policing definitely incorporates the development of profiles for predicting future criminal activity. Concerns about predictive models generate potentially discriminatory predictions therefore are very serious, and need addressing. Potential discrimination may be either overt, though far less likely, or unintended. A valuable case study of which sheds light on such discriminatory data mining practices can be found in US Labour law. It was shown how predictive models could be discriminatory at various stages, from conceptualising the model and training it with training data, to eventually selecting inappropriate features to search for [99]. It is also possible for data scientists to (intentionally or not) use proxies for identifiers like race, income level, health condition and religion. Barocas and Selbst argue that "the current distribution of relevant attributes-attributes that can and should be taken into consideration in apportioning opportunities fairly-are demonstrably correlated with sensitive attributes" [100]. Hence, what may result is unintended discrimination, as predictive models and their subjective and implicit biases are reflected in predicted decisions, or that the discrimination is not even accounted for in the first place. While I have not found any case law where courts have examined such situations in a criminal context, at the very least, law enforcement agencies need to be aware of these possibilities and guard against any forms of discriminatory profiling.

However, Ferguson argues that "the precision of the technology may in fact provide more protection for citizens in broadly defined high crime areas" [101]. This is because the label of a 'high-crime area' may no longer apply to large areas but instead to very specific areas of criminal activity. This implies that previously defined areas of high crime, like entire neighbourhoods may not be scrutinised in such detail. Instead, police now may be more precise in locating and policing areas of high crime, such as an individual street corner or a particular block of flats instead of an entire locality.

Hot Spots

Courts have also considered the existence of notoriously 'high-crime areas as part of considering reasonable suspicion [102]. This was seen in *Illinois v. Wardlow* [103], where the "high crime nature of an area can be considered in evaluating the officer's objective suspicion" [104]. Many cases have since applied this reasoning without scrutinising the predictive value of such a label. In fact, Ferguson asserts that such labelling has questionable evidential value [105]. He uses the facts of the *Wardlow* case itself to challenge the 'high crime area' factor. Ferguson cites the reasoning of one of the judges in the case:

"While the area in question-Chicago's District 11-was a low-income area known for violent crimes, how that information factored into a predictive judgment about a man holding a bag in the afternoon is not immediately clear" [106].

Especially because "the most basic models of predictive policing rely on past crimes" [107], it is likely that the predictive policing methods like hot spot or spatiotemporal analysis and risk terrain modelling may help to gather or build data models about high crime areas. Furthermore, the mathematical rigour of the predictive modelling could help clarify the term 'high crime area'. As Ferguson argues, "courts may no longer need to rely on the generalized high crime area terminology when more particularized and more relevant information is available" [108].

Summary

Ferguson synthesises four themes to which encapsulate reasonable suspicion analysis:

1. Predictive information is not enough on its own. Instead, it is "considered relevant to the totality of circumstances, but must be corroborated by direct police observation" [109].
2. The prediction must also "be particularized to a person, a profile, or a place, in a way that directly connects the suspected crime to the suspected person, profile, or place" [110].
3. It must also be detailed enough to distinguish a person or place from others not the focus of the prediction [111].
4. Finally, predicted information becomes less valuable over time. Hence it must be acted on quickly or be lost [112].

Conclusions from America

The main conclusion to draw from the analysis of the parallels between existing predictions in IV amendment law and predictive policing is that "predictive policing will impact the reasonable suspicion calculus by becoming a factor within the totality of circumstances test" [113]. Naturally, it reaffirms the imperative for predictive techniques to collect reliable data [114] and analyse it transparently [115]. Moreover, in order for courts to evaluate the reliability of the data and the processes used (since predictive methods become part of the reasonable suspicion calculus), courts need to be able to analyse the predictive process. This has implications for the how hearings may be conducted, for how legal adjudicators may require training and many more. Another important concern is that the model of predictive information and police corroboration or direct observation [116] may mean that in areas which were predicted to have low risk of crime, the reasonable suspicion doctrine works against law enforcement. There may be less effort paid to patrolling these other areas as a result of predictions.

Implications for India

While there have been no cases directly involving predictive policing methods, it would be prudent to examine the parts of Indian law which would inform the calculus on the lawfulness of using predictive policing methods. A useful lens to examine this might be found in the observation that prediction is not in itself a novel concept in justice, and is already used by courts and law enforcement in numerous circumstances.

Criminal Procedure in Non-Warrant Contexts

The most logical way to begin analysing the legal implications of predictive policing in India may probably involve identifying parallels between American and Indian criminal procedure, specifically searching for instances where 'reasonable suspicion' or some analogous requirement exists for justifying police searches.

In non-warrant scenarios, we find conditions for officers to conduct such a warrantless search in Section 165 of the Criminal Procedure Code (Cr PC). For clarity purposes I have stated section 165 (1) in full:

"Whenever an officer in charge of a police station or a police officer making an investigation **has reasonable grounds** for believing that anything necessary for the purposes of an

investigation into any offence which he is authorised to investigate may be found in any place with the limits of the police station of which he is in charge, or to which he is attached, and that such thing cannot in his opinion be otherwise obtained without undue delay, such officer may, after recording in writing the grounds of his belief and specifying in such writing, so far as possible, the thing for which search is to be made, search, or cause search to be made, for such thing in any place within the limits of such station." [117]

However, India differs from the USA in that its Cr PC allows for police to arrest individuals without a warrant as well. As observed in *Gulab Chand Upadhyaya vs State Of U.P*, "Section 41 Cr PC gives the power to the police to arrest without warrant in cognizable offences, in cases enumerated in that Section. One such case is of receipt of a 'reasonable complaint' or 'credible information' or 'reasonable suspicion'" [118] Like above, I have stated section 41 (1) and subsection (a) in full:

"41. When police may arrest without warrant.

(1) Any police officer may without an order from a Magistrate and without a warrant, arrest any person-

(a) who has been concerned in any cognizable offence, or against whom a **reasonable complaint has been made, or credible information has been received**, **or a reasonable suspicion exists**, of his having been so concerned" [119]

In analysing the above sections of Indian criminal procedure from a predictive policing angle, one may find both similarities and differences between the proposed American approach and possible Indian approaches to interpreting or incorporating predictive policing evidence.

Similarity of 'reasonable suspicion' requirement

For one, the requirement for "reasonable grounds" or "reasonable suspicion" seems to be analogous to the American doctrine of reasonable suspicion. This suggests that the concepts used in forming reasonable suspicion, for the police to "be able to point to specific and articulable facts which, taken together with rational inferences from those facts, reasonably warrant that intrusion" [120] may also be useful in the Indian context.

One case which sheds light on an Indian interpretation of reasonable suspicion or grounds is *State of Punjab v. Balbir Singh* [121]. In that case, the court observes a requirement for "reason to believe that such an offence under Chapter IV has been committed and, therefore, an arrest or search was necessary as contemplated under these provisions" [122] in the context of Section 41 and 42 in The Narcotic Drugs and Psychotropic Substances Act, 1985 [123]. In examining the requirement of having "reason to believe", the court draws on *Partap Singh (Dr) v. Director of Enforcement, Foreign Exchange Regulation Act* [124], where the judge observed that "the expression 'reason to believe' is not synonymous with subjective satisfaction of the officer. The belief must be held in good faith; it cannot be merely a pretence....." [125]

In light of this, the judge in *Balbir Singh* remarked that "whether there was such reason to believe and whether the officer empowered acted in a bona fide manner, depends upon the facts and circumstances of the case and will have a bearing in appreciation of the evidence" [126]. The standard considered by the court in *Balbir Singh* and *Partap Singh* is different from the 'reasonable suspicion' or 'reasonable grounds' standard as per Section 41 and 165 of Cr PC. But I think the discussion can help to inform our analysis of the idea of reasonableness in law enforcement actions. Of importance was the court requirement of something more than mere "pretence" as well as a belief held in good faith. This could suggest that in fact the reasoning in American jurisprudence about reasonable suspicion might be at least somewhat similar to how Indian courts view reasonable suspicion or grounds in the context of predictive policing, and therefore how we could similarly conjecture that predictive evidence could form part of the reasonable suspicion calculus in India as well.

Difference in judicial treatment of illegally obtained evidence - Indian lack of exclusionary rules

However, the apparent similarity of how police in America and India may act in non-warrant situations - guided by the idea of reasonable suspicion - is only veneered by linguistic parallels. Despite the existence of such conditions which govern the searches without a warrant, I believe that Indian courts currently may provide far less protection against unlawful use of predictive technologies. The main premise behind this argument is that Indian courts refuse to exclude evidence that was obtained in breaches of the conditions of sections of the Cr PC. What exists in place of evidentiary safeguards is a line of cases in which courts routinely admit unlawfully or illegally obtained evidence. Without protections against unlawfully gathered evidence being considered relevant by courts, any regulations on search or conditions to be met before a search is lawful become ineffective. Evidence may simply enter the courtroom through a backdoor.

In the USA, this is by and large, not the case. Although there are exceptions to these rules, exclusionary rules are set out to prevent admission of evidence which violates the constitution [127]. "The exclusionary rule applies to evidence gained from an unreasonable search or seizure in violation of the Fourth Amendment " [128]. Mapp v. Ohio [129] set the precedent for excluding unconstitutionally gathered evidence, where the court ruled that "all evidence obtained by searches and seizures in violation of the Federal Constitution is inadmissible in a criminal trial in a state court" [130].

Any such evidence which then leads law enforcement to collect new information may also be excluded, as part of the "fruit of the poisonous tree" doctrine [131], established in Silverthorne Lumber Co. v. United States [132]. The doctrine is a metaphor which suggests that if the source of certain evidence is tainted, so is 'fruit' or derivatives from that unconstitutional evidence. One such application was in *Beck v. Ohio* [133], where the courts overturned a petitioner's conviction because the evidence used to convict him was obtained via an unlawful arrest.

However in India's context, there is very little protection against the admission and use of unlawfully gathered evidence. In fact, there are a line of cases which lay out the extent of consideration given to unlawfully gathered evidence - both cases that specifically deal with the rules as per the Indian Cr PC as well as cases from other contexts - which follow and develop this line of reasoning of allowing illegally obtained evidence.

One case to pay attention to is *State of Maharastra v. Natwarlal Damodardas Soni* - in this case, the Anti-Corruption Bureau searched the house of the accused after receiving certain information as a tip. The police "had powers under the Code of Criminal Procedure to search and seize this gold if they had reason to believe that a cognizable offence had been committed in respect thereof" [134]. Justice Sarkaria, in delivering his judgement, observed that for argument's sake, even if the search was illegal, "then also, it will not affect the validity of the seizure and further investigation" [135]. The judge drew reasoning from *Radhakishan v. State of U.P* [136]. This which was a case involving a postman who had certain postal items that were undelivered recovered from his house. As the judge in *Radhakishan* noted:

"So far as the alleged illegality of the search is concerned, it is sufficient to say that even assuming that the search was illegal the seizure of the articles is not vitiated. It may be that where the provisions of Sections 103 and 165 of the Code of Criminal Procedure, are contravened the search could be resisted by the person whose premises are sought to be searched. It may also be that because of the illegality of the search the court may be inclined to examine carefully the evidence regarding the seizure. But beyond these two consequences no further consequence ensues." [137]

Shyam Lal Sharma v. State of M.P. [138] was also drawn upon, where it was held that "even if the search is illegal being in contravention with the requirements of Section 165 of the Criminal Procedure Code, 1898, that provision ceases to have any application to the subsequent steps in the investigation" [139].

Even in *Gulab Chand Upadhyay*, mentioned above, the presiding judge contended that even "if arrest is made, it does not require any, much less strong, reasons to be recorded or reported by the police. Thus so long as the information or suspicion of cognizable offence is "reasonable" or "credible", the police officer is not accountable for the discretion of arresting or no arresting" [140].

A more complete articulation of the receptiveness of Indian courts to admit illegally gathered evidence can be seen in the aforementioned *Balbir Singh*. The judgement aimed to:

"dispose of one of the contentions that failure to comply with the provisions of Cr PC in respect of search and seizure even up to that stage would also vitiate the trial. This aspect has been considered in a number of cases and it has been held that the violation of the provisions particularly that of Sections 100, 102, 103 or 165 Cr PC strictly per se does not vitiate the prosecution case. If there is such violation, what the courts have to see is whether any prejudice was caused to the accused and in appreciating the evidence and other relevant factors, the courts should bear in mind that there was such a violation and from that point of view evaluate the evidence on record." [141]

The judges then consulted a series of authorities on the failure to comply with provisions of the Cr PC:

1. *State of Punjab v. Wassan Singh* [142]: "irregularity in a search cannot vitiate the seizure of the articles" [143].
2. *Sunder Singh v. State of U.P* [144]: 'irregularity cannot vitiate the trial unless the accused has been prejudiced by the defect and it is also held that if reliable local witnesses are not available the search would not be vitiated." [145]
3. *Matajog Dobey v.H.C. Bhari* [146]: "when the salutory provisions have not been complied with, it may, however, affect the weight of the evidence in support of the search or may furnish a reason for disbelieving the evidence produced by the prosecution unless the prosecution properly explains such circumstance which made it impossible for it to comply with these provisions." [147]
4. *R v. Sang* [148]: "reiterated the same principle that if evidence was admissible it matters not how it was obtained." [149] Lord Diplock, one of the Lords adjudicating the case, observed that "however much the judge may dislike the way in which a particular piece of evidence was obtained before proceedings were commenced, if it is admissible evidence probative of the accused's guilt "it is no part of his judicial function to exclude it for this reason" [150]. As the judge in *Balbir Singh* quoted from Lord Diplock, a judge "has no discretion to refuse to admit relevant admissible evidence on the ground that it was obtained by improper or unfair means. The court is not concerned with how it was obtained" [151].

The vast body of case law presented above provides observers with a clear image of the courts willingness to admit and consider illegally obtained evidence. The lack of safeguards against admission of unlawful evidence are important from the standpoint of preventing the excessive or unlawful use of predictive policing methods. The affronts to justice and privacy, as well as the risks of profiling, seem to become magnified when law enforcement use predictive methods more than just to augment their policing techniques but to replace some of them. The efficacy and expediency offered by using predictive policing needs to be balanced against the competing interest of ensuring rule of law and due process. In the Indian context, it seems courts sparsely consider this competing interest.

Naturally, weighing in on which approach is better depends on a multitude of criteria like context, practicality, societal norms and many more. It also draws on existing debates in administrative law about the role of courts, which may emphasise protecting individuals and preventing excessive state power (red light theory) or emphasise efficiency in the governing process with courts assisting the state to achieve policy objectives (green light theory) [152].

A practical response may be that India should aim to embrace both elements and balance them appropriately, although what an appropriate balance again may vary. There are some who claim that this balance already exists in India. Evidence for such a claim may come from *R.M. Malkani v. State of Maharashtra* [153], where the court considered whether an illegally tape-recorded conversation could be admissible. In its reasoning, the court drew from *Kuruma, Son of Kanju v. R.* [154], noting that

" if evidence was admissible it matters not how it was obtained. There is of course always a word of caution. It is that the Judge has a discretion to disallow evidence in a criminal case if the strict rules of admissibility would operate unfairly against the accused. That caution is the golden rule in criminal jurisprudence"[155].

While this discretion exists at least principally in India, in practice the cases presented above show that judges rarely exercise that discretion to prevent or bar the admission of illegally obtained evidence or evidence that was obtained in a manner that infringed the provisions governing search or arrest in the Cr PC. Indeed, the concern is that perhaps the necessary safeguards required to keep law enforcement practices, including predictive policing techniques, in check would be better served by a greater focus on reconsidering the legality of unlawfully gathered evidence. If not, evidence which should otherwise be inadmissible may find its way into consideration by existing legal backdoors.

Risk of discriminatory predictive analysis

Regarding the risk of discriminatory profiling, Article 15 of India's Constitution [156] states that "the State shall not discriminate against any citizen on grounds only of religion, race, caste, sex, place of birth or any of them" [157]. The existence of constitutional protection for such forms of discrimination suggests that India will be able to guard against discriminatory predictive policing. However, as mentioned before, predictive analytics often discriminates institutionally, "whereby unconscious implicit biases and inertia within society's institutions account for a large part of the disparate effects observed, rather than intentional choices" [158]. As in most jurisdictions, preventing these forms of discrimination are much harder. Especially in a jurisdiction whose courts are already receptive to allowing admission of illegally obtained evidence, the risk of discriminatory data mining or prejudiced algorithms being used by police becomes magnified. Because the discrimination may be unintentional, it may be even harder for evidence from discriminatory predictive methods to be scrutinised or when applicable, dismissed by the courts.

Conclusion for India

One thing which is eminently clear from the analysis of possible interpretations of predictive evidence is that Indian Courts have had no experience with any predictive policing cases, because the technology itself is still at a nascent stage. There is in fact a long way to go before predictive policing will become used on a scale similar to that of USA for example.

But, even in places where predictive policing is used much more prominently, there is no precedent to observe how courts may view predictive policing. Ferguson's method of locating analogous situations to predictive policing which courts have already considered is one notable approach, but even this does not provide complete answer. One of his main conclusions that predictive policing will affect the reasonable suspicion calculus, or in India's case, contribute to 'reasonable grounds' in some ways, is perhaps the most valid one.

However, what provides more cause for concern in India's context are the limited protections against use of unlawfully gathered evidence. The lack of 'exclusionary rules' unlike those present in the US amplifies the various risks of predictive policing because individuals have little means of redress in such situations where predictive policing may be used unjustly against them.

Yet, the promise of predictive policing remains undeniably attractive for India. The successes predictive policing methods seem to have had In the US and UK coupled with the more

efficient allocation of law enforcement's resources as a consequence of adapting predictive policing evidence this point. The government recognises this and seems to be laying the foundation and basic digital infrastructure required to utilize predictive policing optimally. One ought also to ask whether it is the even within the court's purview to decide what kind of policing methods are to be permissible through evaluating the nature of evidence. There is a case to be made for the legislative arm of the state to provide direction on how predictive policing is to be used in India. Perhaps the law must also evolve with the changes in technology, especially if courts are to scrutinise the predictive policing methods themselves.

ENDNOTES

[1] Joh, Elizabeth E. "Policing by Numbers: Big Data and the Fourth Amendment." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, February 1, 2014. http://papers.ssrn.com/abstract=2403028.

[2] Tene, Omer, and Jules Polonetsky. "Big Data for All: Privacy and User Control in the Age of Analytics." Northwestern Journal of Technology and Intellectual Property 11, no. 5 (April 17, 2013): 239.

[3] Datta, Rajbir Singh. "Predictive Analytics: The Use and Constitutionality of Technology in Combating Homegrown Terrorist Threats." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, May 1, 2013. http://papers.ssrn.com/abstract=2320160.

[4] Johnson, Jeffrey Alan. "Ethics of Data Mining and Predictive Analytics in Higher Education." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, May 8, 2013. http://papers.ssrn.com/abstract=2156058.

[5] Ibid.

[6] Duhigg, Charles. "How Companies Learn Your Secrets." The New York Times, February 16, 2012. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

[7] Ibid.

[8] Lijaya, A, M Pranav, P B Sarath Babu, and V R Nithin. "Predicting Movie Success Based on IMDB Data." International Journal of Data Mining Techniques and Applications 3 (June 2014): 365-68.

[9] Johnson, Jeffrey Alan. "Ethics of Data Mining and Predictive Analytics in Higher Education." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, May 8, 2013. http://papers.ssrn.com/abstract=2156058.

[10] Sangvinatsos, Antonios A. "Explanatory and Predictive Analysis of Corporate Bond Indices Returns." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, June 1, 2005. http://papers.ssrn.com/abstract=891641.

[11] Barocas, Solon, and Andrew D. Selbst. "Big Data's Disparate Impact." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, February 13, 2015. http://papers.ssrn.com/abstract=2477899.

[12] Joh, supra note 1.

[13] US Environmental Protection Agency. "How We Use Data in the Mid-Atlantic Region." US EPA. Accessed November 6, 2015. http://archive.epa.gov/reg3esd1/data/web/html/.

[14] See here for details of blackroom.

[15] Joh, supra note 1, at pg 48.

[16] Perry, Walter L., Brian McInnis, Carter C. Price, Susan Smith and John S. Hollywood. Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations. Santa Monica, CA: RAND Corporation, 2013. http://www.rand.org/pubs/research_reports/RR233. Also available in print form.

[17] Ibid, at pg 2.

[18] Chan, Sewell. "Why Did Crime Fall in New York City?" City Room. Accessed November 6, 2015. http://cityroom.blogs.nytimes.com/2007/08/13/why-did-crime-fall-in-new-york-city/.

[19] Bureau of Justice Assistance. "COMPSTAT: ITS ORIGINS, EVOLUTION, AND FUTURE IN LAW ENFORCEMENT AGENCIES," 2013. http://www.policeforum.org/assets/docs/Free_Online_Documents/Compstat/compstat%20-%20its%20origins%20evolution%20and%20future%20in%20law%20enforcement%20agencies%202013.pdf.

[20] 1996 internal NYPD article "Managing for Results: Building a Police Organization that Dramatically Reduces Crime, Disorder, and Fear."

[21] Bratton, William. "Crime by the Numbers." The New York Times, February 17, 2010. http://www.nytimes.com/2010/02/17/opinion/17bratton.html.

[22] RAND CORP, supra note 16.

[23] RAND CORP, supra note 16, at pg 19.

[24] Joh, supra note 1, at pg 44.

[25] RAND CORP, supra note 16, pg 38.

[26] Ibid.

[27] RAND CORP, supra note 16, at pg 39.

[28] Ibid.

[29] RAND CORP, supra note 16, at pg 41.

[30] Data-Smart City Solutions. "Dr. George Mohler: Mathematician and Crime Fighter." Data-Smart City Solutions, May 8, 2013. http://datasmart.ash.harvard.edu/news/article/dr.-george-mohler-mathematician-and-crime-fighter-166.

[31] RAND CORP, supra note 16, at pg 44.

[32] Joh, supra note 1, at pg 45.

[33] Ouellette, Danielle. "Dispatch - A Hot Spots Experiment: Sacramento Police Department," June 2012. http://cops.usdoj.gov/html/dispatch/06-2012/hot-spots-and-sacramento-pd.asp.

[34] Pitney Bowes Business Insight. "The Safer Derbyshire Partnership." Derbyshire, 2013. http://www.mapinfo.com/wp-content/uploads/2013/05/safer-derbyshire-casestudy.pdf.

[35] Ibid.

[36] Daniel B Neill, Wilpen L. Gorr. "Detecting and Preventing Emerging Epidemics of Crime," 2007.

[37] RAND CORP, supra note 16, at pg 33.

[38] Joh, supra note 1, at pg 46.

[39] Paul, Jeffery S, and Thomas M. Joiner. "Integration of Centralized Intelligence with Geographic Information Systems: A Countywide Initiative." Geography and Public Safety 3, no. 1 (October 2011): 5-7.

[40] Mohler, supra note 30.

[41] Ibid.

[42] Moses, B., Lyria, & Chan, J. (2014). Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools (SSRN Scholarly Paper No. ID 2513564). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2513564

[43] Gorner, Jeremy. "Chicago Police Use Heat List as Strategy to Prevent Violence." Chicago Tribune. August 21, 2013. http://articles.chicagotribune.com/2013-08-21/news/ct-met-heat-list-20130821_1_chicago-police-commander-andrew-papachristos-heat-list.

[44] Stroud, Matt. "The Minority Report: Chicago's New Police Computer Predicts Crimes, but Is It Racist?" The Verge. Accessed November 13, 2015. http://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist.

[45] Moser, Whet. "The Small Social Networks at the Heart of Chicago Violence." Chicago Magazine, December 9, 2013. http://www.chicagomag.com/city-life/December-2013/The-Small-Social-Networks-at-the-Heart-of-Chicago-Violence/.

[46] Lester, Aaron. "Police Clicking into Crimes Using New Software." Boston Globe, March 18, 2013. https://www.bostonglobe.com/business/2013/03/17/police-intelligence-one-click-away/DzzDbrwdiNkjNMA1159ybM/story.html.

[47] Stanley, Jay. "Chicago Police 'Heat List' Renews Old Fears About Government Flagging and Tagging." American Civil Liberties Union, February 25, 2014. https://www.aclu.org/blog/chicago-police-heat-list-renews-old-fears-about-government-flagging-and-tagging.

[48] Rieke, Aaron, David Robinson, and Harlan Yu. "Civil Rights, Big Data, and Our Algorithmic Future," September 2014. https://bigdata.fairness.io/wp-content/uploads/2015/04/2015-04-20-Civil-Rights-Big-Data-and-Our-Algorithmic-Future-v1.2.pdf.

[49] Edmond, Deepu Sebastian. "Jhakhand's Digital Leap." Indian Express, September 15, 2013. http://www.jhpolice.gov.in/news/jhakhands-digital-leap-indian-express-15092013-18219-1379316969.

[50] Jharkhand Police. "Jharkhand Police IT Vision 2020 - Effective Shared Open E-Governance." 2012. http://jhpolice.gov.in/vision2020. See slide 2

[51] Edmond, supra note 49.

[52] Edmond, supra note 49.

[53] Kumar, Raj. "Enter, the Future of Policing - Cops to Team up with IIM Analysts to Predict & Prevent Incidents." The Telegraph. August 28, 2012. http://www.telegraphindia.com/1120828/jsp/jharkhand/story_15905662.jsp#.VkXwxvnhDWK.

[54] Ibid.

[55] Ibid.

[56] Ibid.

[57] See supra note 49.

[58] See here: http://dashboard.jhpolice.gov.in/ for Jharkhand Police crime dashboard.

[59] Lavanya Gupta, and Selva Priya. "Predicting Crime Rates for Predictive Policing." Gandhian Young Technological Innovation Award, December 29, 2014. http://gyti.techpedia.in/project-detail/predicting-crime-rates-for-predictive-policing/3545.

[60] Gupta, Lavanya. "Minority Report: Minority Report." Accessed November 13, 2015. http://cmuws2014.blogspot.in/2015/01/minority-report.html.

[61] See supra note 59.

[62] See here: http://bprd.nic.in/showfile.asp?lid=1224 for details about 44th All India Police Science Congress.

[63] India, Press Trust of. "Police Science Congress in Gujarat to Have DRDO Exhibition." Business Standard India, March 10, 2015. http://www.business-standard.com/article/pti-stories/police-science-congress-in-gujarat-to-have-drdo-exhibition-115031001310_1.html.

[64] National Crime Records Bureau. "About Crime and Criminal Tracking Network & Systems - CCTNS." Accessed November 13, 2015. http://ncrb.gov.in/cctns.htm.

[65] Ibid. (See index page)

[66] U.S. Const. amend. IV, available here: https://www.law.cornell.edu/constitution/fourth_amendment

[67] United States v Katz, 389 U.S. 347 (1967) , see here: https://supreme.justia.com/cases/federal/us/389/347/case.html

[68] See supra note 1, at pg 60.

[69] See supra note 1, at pg 60.

[70] Villasenor, John. "What You Need to Know about the Third-Party Doctrine." The Atlantic, December 30, 2013. http://www.theatlantic.com/technology/archive/2013/12/what-you-need-to-know-about-the-third-party-doctrine/282721/.

[71] Smith v Maryland, 442 U.S. 735 (1979), see here: https://supreme.justia.com/cases/federal/us/442/735/case.html

[72] United States v Jones, 565 U.S. ___ (2012), see here: https://supreme.justia.com/cases/federal/us/565/10-1259/

[73] Newell, Bryce Clayton. "Local Law Enforcement Jumps on the Big Data Bandwagon: Automated License Plate Recognition Systems, Information Privacy, and Access to Government Information." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, October 16, 2013. http://papers.ssrn.com/abstract=2341182, at pg 24.

[74] See supra note 72.

[75] Dahyabhai Chhaganbhai Thakker vs State Of Gujarat, 1964 AIR 1563

[76] See supra note 16.

[77] See supra note 66.

[78] Brinegar v. United States, 338 U.S. 160 (1949), see here: https://supreme.justia.com/cases/federal/us/338/160/case.html

[79] Terry v. Ohio, 392 U.S. 1 (1968), see here: https://supreme.justia.com/cases/federal/us/392/1/case.html

[80] Ferguson, Andrew Guthrie. "Big Data and Predictive Reasonable Suspicion." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, April 4, 2014. http://papers.ssrn.com/abstract=2394683, at pg 287. See also supra note 79.

[81] See supra note 80.

[82] See supra note 80.

[83] See supra note 80.

[84] See supra note 80, at pg 289.

[85] Illinois v. Gates, 462 U.S. 213 (1983). See here: https://supreme.justia.com/cases/federal/us/462/213/case.html

[86] See Alabama v. White, 496 U.S. 325 (1990). See here: https://supreme.justia.com/cases/federal/us/496/325/

[87] See supra note 80, at pg 291.

[88] See supra note 80, at pg 293.

[89] See supra note 80, at pg 308.

[90] Ibid.

[91] Ibid.

[92] Larissa Cespedes-Yaffar, Shayona Dhanak, and Amy Stephenson. “U.S. v. Mendenhall, U.S. v. Sokolow, and the Drug Courier Profile Evidence Controversy.” Accessed July 6, 2015. http://courses2.cit.cornell.edu/sociallaw/student_projects/drugcourier.html.

[93] Ibid.

[94] United States v. Sokolow, 490 U.S. 1 (1989), see here: https://supreme.justia.com/cases/federal/us/490/1/

[95] See supra note 80, at pg 295.

[96] See supra note 80, at pg 297.

[97] See supra note 80, at pg 308.

[98] See supra note 80, at pg 310.

[99] See supra note 11.

[100] See supra note 11.

[101] See supra note 80, at pg 303.

[102] See supra note 80, at pg 300.

[103] Illinois v. Wardlow, 528 U.S. 119 (2000), see here: https://supreme.justia.com/cases/federal/us/528/119/case.html

[104] Ibid.

[105] See supra note 80, at pg 301.

[106] Ibid.

[107] See supra note 1, at pg 42.

[108] See supra note 80, at pg 303.

[109] See supra note 80, at pg 303.

[110] Ibid.

[111] Ibid.

[112] Ibid.

[113] See supra note 80, at pg 312.

[114] See supra note 80, at pg 317.

[115] See supra note 80, at pg 319.

[116] See supra note 80, at pg 321.

[117] Section 165 Indian Criminal Procedure Code, see here: https://indiankanoon.org/doc/996365/

[118] Gulab Chand Upadhyaya vs State Of U.P, 2002 CriLJ 2907

[119] Section 41 Indian Criminal Procedure Code

[120] See supra note 79

[121] State of Punjab v. Balbir Singh. (1994) 3 SCC 299

[122] Ibid.

[123] Section 41 and 42 in The Narcotic Drugs and Psychotropic Substances Act 1985, see here: https://indiankanoon.org/doc/1727139/

[124] Partap Singh (Dr) v. Director of Enforcement, Foreign Exchange Regulation Act. (1985) 3 SCC 72 : 1985 SCC (Cri) 312 : 1985 SCC (Tax) 352 : AIR 1985 SC 989

[125] Ibid, at SCC pg 77-78.

[126] See supra note 121, at pg 313.

[127] Carlson, Mr David. “Exclusionary Rule.” LII / Legal Information Institute, June 10, 2009. https://www.law.cornell.edu/wex/exclusionary_rule.

[128] Ibid.

[129] Mapp v Ohio, 367 U.S. 643 (1961), see here: https://supreme.justia.com/cases/federal/us/367/643/case.html

[130] Ibid.

[131] Busby, John C. “Fruit of the Poisonous Tree.” LII / Legal Information Institute, September 21, 2009. https://www.law.cornell.edu/wex/fruit_of_the_poisonous_tree.

[132] Silverthorne Lumber Co., Inc. v. United States, 251 U.S. 385 (1920), see here: https://supreme.justia.com/cases/federal/us/251/385/case.html

[133] Beck v. Ohio, 379 U.S. 89 (1964), see here: https://supreme.justia.com/cases/federal/us/379/89/case.html

[134] State of Maharashtra v. Natwarlal Damodardas Soni, (1980) 4 SCC 669, at 673.

[135] Ibid.

[136] Radhakishan v. State of U.P. [AIR 1963 SC 822 : 1963 Supp 1 SCR 408, 411, 412 : (1963) 1 Cri LJ 809]

[137] Ibid, at SCR pg 411-12.

[138] Shyam Lal Sharma v. State of M.P. (1972) 1 SCC 764 : 1974 SCC (Cri) 470 : AIR 1972 SC 886

[139] See supra note 135, at page 674.

[140] See supra note 119, at para. 10.

[141] See supra note 121, at pg 309.

[142] State of Punjab v. Wassan Singh, (1981) 2 SCC 1 : 1981 SCC (Cri) 292

[143] See supra note 121, at pg 309.

[144] Sunder Singh v. State of U.P, AIR 1956 SC 411 : 1956 Cri LJ 801

[145] See supra note 121, at pg 309.

[146] Matajog Dobey v.H.C. Bhari, AIR 1956 SC 44 : (1955) 2 SCR 925 : 1956 Cri LJ 140

[147] See supra note 121, at pg 309.

[148] R v. Sang, (1979) 2 All ER 1222, 1230-31

[149] See supra note 121, at pg 309.

[150] Ibid.

[151] Ibid.

[152] Harlow, Carol, and Richard Rawlings. Law and Administration. 3rd ed. Law in Context. Cambridge University Press, 2009.

[153] R.M. Malkani v. State of Maharashtra, (1973) 1 SCC 471

[154] Kuruma, Son of Kanju v. R., (1955) AC 197

[155] See supra note 154, at 477.

[156] Indian Const. Art 15, see here: https://indiankanoon.org/doc/609295/

[157] Ibid.

[158] See supra note 11.

Digital Emergency Power: Big Data and Ebola Response in Liberia

SEAN MARTIN MCDONALD

Introduction

Digitizing disaster response invites the problems of digital systems into the most fragile and vulnerable environments in the world. Troublingly, it is often humanitarian organizations that lead the charge, underestimating the practical and legal implications of digitizing these systems, from data security to operational coordination to the fairness of algorithms. In addition to their own digital transformations, many humanitarian organizations actively encourage governments, charitable foundations, technology companies, and mobile networks to share data in ways that are illegal without user consent or the invocation of governmental emergency powers. The governance of emergency powers over digital systems remain poorly defined and badly regulated, and lack the basic due process checks and balances that exist for nearly every other kind of government emergency authority. The humanitarian community knows that it does not have the technological, legal, or institutional checks necessary to fairly or fully realize the promise of digital systems. That knowledge, however, hasn't prevented many of the world's most important and trusted institutions from taking irresponsible, at best, and illegal, at worst, risks with some of the world's most sensitive data.

The most prominent of these risks is the growing call for mobile network operators – the companies that provide mobile phone and data services – databases, which usually contain a significant amount of personal information. According to the most frequently cited expert for justifying the release of mobile phone records, we simply don't have a sufficient understanding of how to apply these records to social services systems—even less so in fragile contexts. The humanitarian community lacks the data modeling, professional technology implementation standards, and the enforcement capacity to protect, or even define, the public's interest. Even in places where there are regulatory institutions to fill the gap, emergency contexts often cause disruption or suspension of the safeguarding processes required to protect human rights. Without these processes – or any other form of public oversight – digitizing humanitarian systems can add layers of opacity to the already complex data models, implementation approaches, or intended outcomes of the response, further crippling tenuous public trust and good governance.

Where this leaves us is a world where the stretching and violation of national, regional, and international human rights and data protection laws has become the norm, for a benefit that practically eludes definition, if it even exists. Despite that, many governments, businesses, and international organizations are routinely given access to mobile network operator data – Call Detail Records (CDRs), containing some of the most personal, re-identifiable data that people produce today. Humanitarian, academic, and journalistic calls for the release of CDRs gain the most traction in emergencies, contributing to a norm that actively disregards individual rights and consent. These practices not only put individuals at risk of harm, but as digital jurisprudence continues to come into its own, they will put many of the world's best—and best-intentioned—international organizations at risk of onerous lawsuits...

In 2014, a small outbreak of Ebola in West Africa grew into a global health crisis. Over the ensuing 18 months, Ebola infected almost 30,000 people and killed over 11,000. Although a large number of factors led to the spread of the disease, the early stages were characterized by rumor, misinformation, and a deep distrust of the institutions responding to the disaster. By the time the international community mobilized, the epidemic had been metastasizing for nine months, and required significantly more investment than initially anticipated. International health and humanitarian response agencies made apocalyptic predictions

about the epidemic, sparking panic and uncertainty. A new wave of foundations, donors, and organizations took an interest in the response, leading to a surge in people, resources, and operational chaos. That chaos fed a growing narrative that the problem in the response effort was a lack of good information technology and, more specifically, data.

CDRs and Contact Tracing Ebola

CDRs include some of the most sensitive data that people generate, including location history, substantive communication history, and personal billing data. Despite that, there is very little evidence to suggest that CDRs, especially those that have been anonymized, are useful to track the spread of the Ebola virus. Nevertheless, the farthest-reaching, most common, and least interrogated requests for data during the epidemic were the international community's push for the release of CDRs to aid response efforts. This case study performs an in-depth investigation of the way that CDRs could be used to track Ebola, revealing that they're only useful when re-identified, invalidating anonymization as a balancing approach to privacy, and thus legal, protection. Not only are CDRs an ineffective way to track the Ebola virus, sharing CDR data likely violates data protection law.

In August of 2014, the MIT Technology Review published an article suggesting that mobile network data could be used to predict the spread of the Ebola virus [1]. The article drew on the work of Flowminder, a Swedish data science non-profit, and Caroline Buckee, a Harvard epidemiologist [2]. The idea and concept grew substantially from there – both in media coverage, complementary academic studies, and among development donors – leading to numerous requests for mobile networks and the Liberian Telecommunications Authority (LTA) to release CDRs. The most typical version of the request was that CDRs could be anonymized to reflect aggregated population movement, which could also be used to build predictive models about the spread of Ebola. More aggressive versions of those requests suggested that CDRs could be used to perform contact tracing to track the infected – and so mobile networks should open access to their core user databases to donor organizations. Over the course of September and October, coverage and pressure grew, and ultimately networks in Sierra Leone and Guinea did share CDRs...

Contact tracing is one of the most aggressive and manual ways that health systems track a disease. In its most basic form, it identifies an infected person and interviews them to track every person that they've been in contact with during the period they were contagious. Historically, it's an analogue process involving significant time from dedicated teams who interview the infected, visit the locations, interview and test others who have been exposed, and then repeat the process, to interview everyone the second person has been in touch with (provided they're positive for the virus).

Contact tracing is particularly effective for diseases that require direct contact with a contagious source in order to spread. Ebola is a hemorrhagic fever that's manually transmitted through bodily fluids, meaning that it is not as contagious as many other diseases. However, the virus causes victims to expel a wide range of fluids, meaning that contact in later stages of infection is extremely contagious, even after death. Ebola requires direct contact with the fluids of an infected person after they've begun exhibiting symptoms. Despite having a high mortality rate, Ebola is relatively difficult to transmit. For context, the R0 scale – the definitive system used to rate how quickly a disease spreads – gave the Ebola epidemic in West Africa a 1.51-2.53 rating, varying by country [3]. That means that each person infected was likely to infect an average of 2 additional people. For comparison, Measles has a R0 score of 18 [4]. The rate at which Ebola spreads is important, but less important for this analysis than how it spreads. The way Ebola spreads determines what data points are useful predictors of additional infections.

During the Ebola epidemic, the international response put a significant focus on contact tracing. Like nearly every other aspect of the epidemic, the teams initially used paper-based records, realized the challenges involved, and transitioned to digital processes. Centrally, the WHO and the MoH brought in Dr. Hans Rosling, an international expert on data science and

processes. Dr. Rosling built a small team to manually digitize the paper records that were collected from the Ebola Treatment Units (ETU), catalogue and then held daily briefings on the state of the epidemic [5]. Eventually, the county and district levels developed their own Data Officers, who ensured reporting and digitization into decentralized databases and then aggregated them into the national system [6]...

The other major digitization effort in the contact tracing practice focused on contact tracers – namely, giving contact tracers digital data entry tools so that the information collected was digital and portable from the outset. Given the large number of mobile and digital data collection tools, this particular opportunity was more sought-after and contested. According to Bawo, there were at least 9 separate contact tracing applications pitched to him – and he wasn't primarily responsible for that vertical of work [7]. The Liberian and international response community mostly responded to these pitches the same way they reacted to others – namely that they had "good enough" systems in place, and with few examples didn't adopt new tools after the peak of the epidemic. In November of 2014, Paul Allen organized the donation of 10,000 mobile phones, which were distributed to contact tracing teams, and added value – though providing scratch cards (the vehicle for delivering air time and messaging credit) was an ongoing challenge throughout the response [8]. That said, digitizing the existing work of the contact tracers, once implemented, made a sizable difference to the speed and efficiency of digitization.

This was the context for the call for CDR data...

Flowminder's migration analysis using CDRs is the most broadly referenced justification for releasing CDRs to trace and predict the spread of Ebola, though the organization's primary assertion is that real-time mobility data would have been valuable as a means of building targeted and adaptable containment strategies [9]. While it's possible – and certainly applicable in emergencies with significant variations in movement patterns – there's a significant risk of overestimating the potential benefits of dynamic migration analysis of the transmission of this virus, held against the very real privacy costs of compelled CDR release. Even Dr. Linus Bengtsson, the Executive Director of Flowminder, is careful to nuance the significant way that virus transmission affects the importance of migration data to the data model:

If you're planning traffic, you need the whole picture to model the road. If you're trying to catch an individual traffic offender, the model of traffic can help – but it's really an entirely different ball game... With Ebola, you need to see the other person, which you don't see in the mobile data, so it's not as interesting [for Ebola] as for vector-born diseases... Just because you go to the same 100km/sq. area doesn't mean you met someone [10].

There are personal identifiers and cultural behaviors that are more likely to correlate to the probability of the virus's spread, such as burial practices, which have a measurable impact on post-mortem transmission. Those datasets are often unavailable or highly dependent on correlation to other datasets in order to provide relevant insight (for example, maps of prevalence of cultural behaviors exist, but not in ways that directly document burial practices and the degree of direct physical contact involved in each). That use of burial practice information, however, clearly falls within the definitions of personally identifiable information – undermining assertions of the independent value of anonymized data in predicting the spread of Ebola. In other words, CDRs are only useful to Ebola contact tracing when they're re-identified or linked to personally identifiable information. As a question of data modeling, that increases the benefit of using a dataset like CDRs, as there aren't many other data sources that contain real-time relevant information for predictive modeling.

Ultimately, this simplifies the analysis of privacy, because it eliminates the presumption that anonymized data is useful to track or predict Ebola. There is no reason to believe that there are technological ways to ensure privacy in CDRs – both because of the necessary correlations and the significant risks of re-identification. As a Brookings Institution white paper on mobile data collection in disease tracking said, "Best practices should accept

that there are no perfect ways to de-identify data and there probably never will be" [11]. Recognizing that – whether by necessity or technological insecurity – CDRs are only useful to track Ebola as personally identifiable datasets. Dispensing with the fiction that anonymization is more than intermediary protection for data sharers, at least for the purpose of containing a contact-based virus, significantly simplifies the ethical and legal analysis...

Digital Emergency Power and Legal Responses

The chaos of humanitarian disaster often creates an implied social license for experimentation with new approaches, under the assumption of better outcomes. Vested interests dominate the public discussion of humanitarian data modeling, downplaying the dangers of what is essentially a public experiment to combine mobile network data and social engineering algorithms. In the case of using mobile network data to track or respond to Ebola, the approaches are so new–and generally so illegal– that most advocacy focuses on securing basic access to data. Advocates for the release of CDRs often paint an optimistic picture of its potential benefits, without applying the same rigor to the risks or likelihood of harm. This trades on the social license created by disaster to experiment with the lives of those affected, under the implicit assumption that it can't make the situation worse.

The problem is that personally identifiable data in the wrong hands can make many situations worse. Once released, data is nearly impossible to control or "take back", meaning that once released CDRs create risks immediately that last indefinitely. During and after a disaster, there are substantial benefits to having access to CDRs in humanitarian and commercial contexts. The value of these data sets, and the rights of the people they represent, should increase the ethical concerns and scrutiny of access negotiated under the auspices of humanitarian and development contexts. In the Ebola response, it did just the opposite.

Data modeling may have significant benefits, but most publicly funded social interventions bear the burden of proving those benefits before compromising the human rights of millions of people. To date, there simply hasn't been nuanced enough experimentation, analysis, or debate about what factors lead to data modeling having a positive social impact that addresses specific threats or issues. The debate about the humanitarian use of data models not only has that burden not been met, there aren't even the appropriate research, licensing, and oversight mechanisms to evaluate and protect against potential harms. Without more research, the forced release of mobile network data may not only endanger the lives of the people humanitarians seek to serve – they may expose humanitarian organizations to a new kind of legal risk.

Calls for CDRs seems to gain the most traction during times of emergency, which also has the potential to set norms in circumstances that distort the cost-benefit analysis away from individual rights and agency. This chapter will explore the state of the institutional debate, in order to understand the incentives and actors involved in defining the legal and practical exercise of digital emergency powers...

Although there are a significant number of ways to structure the release of CDRs, most of the public calls for openness have been direct appeals to mobile network operators and governments. In both cases, releasing personally identifiable information absent customer consent is an extra-legal action, requiring governmental sanction. During times of emergency, the government is typically the authority compelling the release or receiving CDR data. Governments can share or delegate information and assets to third parties, but usually only within limitations that attach public accountability to that delegation. Compelling the release of CDRs, no matter the purpose or use, is an exercise of emergency power – which invokes a wide range of legal issues, especially for third parties that act as quasi-governmental bodies.

Calls for CDRs, even when they recognize the experimental nature of mobile data in social and public health systems, gloss over the significance of the inherent costs of facilitating

that experimentation. Specifically, releasing CDRs infringes on the property and human rights of the people who generate, and therefore own, the underlying data. In addition to ownership and privacy rights, the release of personally identifiable data creates a significant and indefinite amount of risk to the individuals affected. At the moment, however, the only way to seek protection is through the legal system. The legal systems and regulatory authorities charged with monitoring and enforcing data protection laws– even when they have implementation frameworks– continue to struggle with the technological complexity and significant costs of enforcement of licensing personally identifiable information...

For the most part, how the humanitarian community acquires, moves, uses, stores, shares, and publishes data is likely illegal. The non-governmental use of CDRs is so illegal that most writing advocating for their use cite the law as a major obstacle for continued experimentation. Awareness does not mitigate illegality, however. The most commonly discussed legal risks draw from telecommunications, privacy, and data protection laws, but current humanitarian practice could also implicate property, libel, and due process laws. The differences and conflicts between these laws run the risk of compounding, as opposed to mitigating, the liability faced by international humanitarian organizations. An international organization can be brought before courts, tribunals, and regulatory authorities anywhere they operate–and conviction in one jurisdiction doesn't protect that same organization from multiple suits for the same case or practice.

The most common defense humanitarian data advocates offer is that CDRs are anonymized prior to sharing, and therefore aren't subject to data protection laws. However, the impossibility of complete anonymization, particularly as more data is released, belies this argument. Commercial telecommunication regulations and data protection laws, prohibit the sharing or seizure of personally identifiable information with or by any party other than the government, especially absent licensing and enforcement mechanisms to protect those affected. While these legal protections are inconvenient for organizations seeking to experiment with emergency response contexts, they do play an important role in the protection of human and property rights...

Humanitarian organizations are likely subject to an enormous range of laws, regulatory regimes, and digital authorities outside of Liberia and ECOWAS that create additional causes of action, depending on the circumstances and affected parties. For example, children have heightened privacy protections in international law – meaning that any organization focusing on children may be subject to heightened scrutiny or consequence [12]. Similarly, a number of legal entities – including the United States and European Union – have laws that govern the way that organizations registered within their jurisdictions treat civil tort liability, data privacy, and human rights wherever they operate. As a result international organizations and officers that receive or user of data illegally may be subject to litigation in their country of origin, as well as the country where the data was acquired. Some of those entities also have long-arm jurisdictional laws, such as the Alien Tort Claims Act, which could be applied to bring suit against mobile network operators that violate the human rights of their citizens, as enunciated in international law.

These long-arm lawsuits are most likely in cases where the release, processing, or use of CDRs is alleged as a potential or partial cause of damages to a person's health, freedoms, or identity.

There is also a rapidly evolving body of law that regulates the transfer of data across national or regional boundaries, which may form the foundation of civil causes of action in a wide range of sovereign jurisdictions and non-governmental fora. Similarly, academic institutions have their own standards for treatment of personally identifying information inline with their own ethical and Internal Review Boards, which operate independently – and create separate causes of action within their designated adjudication systems...

Although this is not the first time that disaster response operations have been used to justify investment in untested approaches, it is one of the most invasive. Unlike many other forms of experimentation that happen during emergency responses, data lasts forever. It is

also almost impossible to track or control once released. Not only is it premature to ascribe value to the abrogation of the privacy of millions of people, it's impossible to how far or for how long those datasets may pose a risk to the people they identify. There simply aren't yet strong enough global technological, operational, organizational, legal, or economic mechanisms in place to measure the benefits or manage the risks involved.

The laws that do exist to govern this type of data request suggest that this kind of expropriation, undertaken without any form of due process, would be illegal. Given the novelty of the request, there's no precedent for this type of data seizure, nor is there any certainty that the available enforcement mechanisms have the capacity or the requisite independence to evaluate the costs or benefits of such a practice. What is certain is the growing interest and willingness of a wide range of actors to take advantage of these datasets during the periods of panic surrounding emergency – often in ways that limit fundamental freedoms with the threat of force.

As so often happens when confronted with potential – especially in emergency – the international community underestimated the present and future risks of using CDRs and data modeling to guide disaster response. There is little question that the use of appropriate data to feed the operations of a well-coordinated response effort has significant potential to reduce human suffering. But that potential is far from realized, and if we're going to realize it, we'll need to build the institutional coordination, mathematical validation, and legal rights frameworks so that the benefits of digitization evolve alongside our ability to control them.

ENDNOTES

[1] Talbot, David "Cell-Phone Data Might Help Predict Ebola's Spread." MIT Technology Review (online) 22 August 2014: http://www.technologyreview.com/news/530296/cell-phone-data-might-help-predict-ebolas-spread/.

[2] Ibid.

[3] Althaus, Christian L.; "Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa," PLOS (online) 2 September 2014: http://currents.plos.org/outbreaks/article/estimating-the-reproduction-number-of-zaire-ebolavirus-ebov-during-the-2014-outbreak-in-west-africa/.

[4] Britton, Tom; Diekmann, Odo; Heesterbeek, Hans; "Mathematical Tools for Understanding Infectious Disease Dynamics," Princeton University Press, 2013 Chapter 7, pg. 161

[5] Interview, Dr. Ling Kituyi, Foreign Medical Team Coordinator, World Health Organization. 18 September 2015.

[6] Interview. Augustine Kornyon, Data Coordinator for the International Rescue Committee. 24 September 2015.

[7] Interview, Luke Bawo, Head of Information Management Systems, Ministry of Health. 26 September 2015

[8] "Paul G. Allen Commits to Enhancing Communication Capabilities in West Africa to Help Fight Ebola." Paul G. Allen Foundation, News (online) 17 November 2014: http://www.pgafamilyfoundation.org/news/news-articles/press-releases/enhancing-communication-to-fight-ebola. Even the Emergency Dispatch Unit (EDU), which manages the national emergency help line and routes ambulances and investigation teams, faces an ongoing shortage of scratch cards – they are temporarily provided by the International Committee of the Red Cross (ICRC), but that program is scheduled to stop operations in November of 2015. Interview, Dr. Ling Kituyi, Foreign Medical Team Coordinator, World Health Organization. 18 September 2015. Interview, Arthur Vaye, Head of the Emergency Dispatch Unit, 23 September 2015.

[9] Bengtssson, Linus; Buckee, Caroline O.; Lu, Xin; Tatem, Andrew J.; Wetter, Erik; and Wesolowski, Amy. "Commentary: Containing the Ebola Outbreak – the Potential and Challenge of Mobile Network Data." PLOS (online) 29 September 2014: http://currents.plos.org/outbreaks/article/containing-the-ebola-outbreak-the-potential-and-challenge-of-mobile-network-data/.

[10] Interview, Dr. Linus Bengtsson, Executive Director of Flowminder. 11 September 2015.

[11] Kendall, Jake; Kerry, Cameron F.; and Montjoye, Alexandre de. "Enabling Humanitarian use of Mobile Phone Data," Issues in Technology Innovation (online) November 2014: http://www.brookings.edu/~/media/research/files/papers/2014/11/12-enabling-humanitarian-mobile-phone-data/brookingstechmobilephonedataweb.pdf.

[12] These rights broadly emanate from the Universal Declaration of Human Rights, and its subsequent amendments and interpretations, including the UN Convention on the rights of the child.

Are We Throwing Our Data Protection Regimes Under the Bus?

ROHAN GEORGE

Consent is complicated. What we think of as reasonably obtained consent varies substantially with the circumstance. For example, in treating rape cases, the UK justice system has moved to recognise complications like alcohol and its effect on explicit consent [1]. Yet in contracts, consent may be implied simply when one person accepts another's work on a contract without objections [2]. These situations highlight the differences between the various forms of informed consent and the implications on its validity.

Consent has emerged as a key principle in regulating the use of personal data, and different countries have adopted different regimes, ranging from the comprehensive regimes like of the EU to more sectoral approaches like that in the USA. However, in our modern epoch characterised by the big data analytics that are now commonplace, many commentators have challenged the efficacy and relevance of consent in data protection. I argue that we may even risk throwing our data protection regimes under the proverbial bus should we continue to focus on consent as a key pillar of data protection.

Consent as a Tool in Data Protection Regimes

In fact, even a cursory review of current data protection laws around the world shows the extent of the law's reliance on consent. In the EU for example, Article 7 of the Data Protection Directive, passed in 1995, provides that data processing is only legitimate when "the data subject has unambiguously given his consent" [3]. Article 8, which guards against processing of sensitive data, provides that such prohibitions may be lifted when "the data subject has given his explicit consent to the processing of those data" [4]. Even as the EU attempts to strengthen data protection within the bloc with the proposed reforms to data protection [5], the focus on the consent of data subject remains strong. There are proposals for an "unambiguous consent by the data subject" [6] requirement to be put in place. Such consent will be mandatory before any data processing can occur [7].

Despite adopting very different overall approaches to data protection and privacy, consent is an equally integral part of data protection frameworks in the USA. In his book Protectors of Privacy [8], Abraham Newman describes two main types of privacy legislation: comprehensive and limited. He argues that places like the EU have adopted comprehensive regimes, which primarily seek to protect individuals because of the "informational and power asymmetry" between individuals and organisations [9]. On the other hand, he classifies the American approach as limited, focusing on more sectoral protections and principles of fair information practice instead of overarching legislation [10]. These sectors include the Fair Credit Reporting Act [11] (which governs consumer credit reporting), the Privacy Act [12] (which governs data collected by Federal government) and Electronic Communications Privacy Act [13] (which deals with email communications) among others. However, the Federal Trade Commission describes itself as having only "limited authority over the collection and dissemination of personal data collected online" [14].

This is because the general data processing that is commonplace in today's era of big data is only regulated by the privacy protections that come from the Federal Trade Commission's (FTC) Fair Information Practice Principles (FIPPs). Expectedly, consent is equally important under the FTC's FIPPs. The FTC describes the principle of consent as "the second widely-accepted core principle of fair information practice" [15] in addition to the principle of notice. Other guidelines on fair data processing published by organisations like the Organisation for Economic Cooperation and Development [16] (OECD) or Canadian Standards Association [17] (CSA) also include consent as a key mechanism in data protection.

The Origins of Consent in Privacy and Data Protection

Given the clearly extensive reliance on consent in data protection, it seems prudent to examine the origins of consent in privacy and data protection. Just why does consent have so much weight in data protection?

One reason is that data protection, along with inextricably linked concerns about privacy, could be said to be rooted in protecting private property. It was argued that the "early parameters of what was to become the right to privacy were set in cases dealing with unconventional property claims" [18], such as unconsented publication of personal letters [19] or photographs [20]. It was the publication of Brandeis and Warren's well-known article "The Right to Privacy" [21], that developed "the current philosophical dichotomy between privacy and property rights" [22], as they asserted that privacy protections ought to be recognised as a right in and of themselves and needed separate protection [23]. Indeed, it was Warren and Brandeis who famously borrowed Justice Cooley's expression that privacy is the "right to be let alone" [24].

On the other side of the debate are scholars like Epstein and Posner, who see privacy protections as part of protecting personal property under tort law [25]. However, the central point is that most scholars seem to acknowledge the relationship between privacy and private property. Even Brandeis and Warren themselves argued that one general aim of privacy is "to protect the privacy of private life, and to whatever degree and in whatever connection a man's life has ceased to be private" [26].

It is also important to locate the idea of consent within the domain of privacy and private property protections. Ostensibly, consent seems to have the effect of lessening the privacy protections afforded in a particular situation to a person, because by acquiescing to the situation, one could be seen as waiving their privacy concerns. Brandeis and Warren concur with this position as they acknowledge how "the right to privacy ceases upon the publication of the facts by the individual, or with his consent" [27]. They assert that this is "but another application of the rule which has become familiar in the law of literary and artistic property" [28].

Perhaps the most eloquent articulation of the importance of consent in privacy comes from Sir Edward Coke's idea that "every man's house is his castle" [29]. Though the 'Castle Doctrine' has been used as a justification for protecting one's property with the use of force [30], I think that implied in the idea of the 'Castle Doctrine' is that consent is necessary in order to preserve privacy. If not, why would anyone be justified in preventing trespass, other than to prevent unconsented entry or use of their property. The doctrine of "Volenti non fit injuria" [31], or 'to one who consents no injury is done', is thus the very embodiment of the role of consent in protecting private property. And as conceptions of private property develop to recognise that the data one gives out is part of his private property, for example in *US v. Jones*, which led scholars to assert that "people should be able to maintain reasonable expectations of privacy in some information voluntarily disclosed to third parties" [32], so does consent act as an important aspect of privacy protection.

Yet, linking privacy with private property is not universally accepted as the conception of privacy. For instance, Alan Westin, in his book Privacy and Freedom [33], describes privacy as "the right to control information about oneself" [34]. Another scholar, Ruth Gavison, contends instead that "our interest in privacy is related to our concern over our accessibility to others: the extent to which we are known to others, the extent to which others have physical access to us, and the extent to which we are the subject of others' attention" [35].

While these alternative notions about privacy's foundational principles may differ from those related to linking privacy with private property, locating consent within these formulations of privacy is possible. Regarding Westin's argument, I think that implicit in the right to control one's information are ideas about individual autonomy, which is exercised through giving or withholding one's consent. Similarly, Gavison herself states that privacy functions to advance "liberty, autonomy and selfhood" [36]. Consent plays a key role in upholding this liberty,

autonomy and selfhood that privacy affords us. Clearly therefore, it is far from unfounded to claim that consent is an integral part of protecting privacy.

Consent, Big Data and Data Protection

Given the solid underpinnings of the principle of consent in privacy protection, it was hardly a coincidence that consent became an integral part of data protection. However, with the rise of big data practices, one quickly finds that consent ceases to work effectively as a tool for protecting privacy. In a big data context, Solove argues that privacy regulation rooted in consent is ineffective, because garnering consent amidst ubiquitous data collection for all the online services one uses as part of daily life is unmanageable [37]. Additionally, the secondary uses of one's data are difficult to assess at the point of collection, and subsequently meaningful consent for secondary use is difficult to obtain [38]. This section examines these two primary consequences of prioritising consent amidst Big data practises.

Consent Places Unrealistic and Unfair Expectations on the Individual

As noted by Tene and Polonetsky, the first concern is that current privacy frameworks which emphasize informed consent "impose significant, sometimes unrealistic, obligations on both organizations and individuals" [39]. The premise behind this argument stems from the way that consent is often garnered by organisations, especially regarding use of their services. An examination of various terms of use policies from banks, online video streaming websites, social networking sites, online fashion or more general online shopping websites reveals a deluge of information that the user has to comprehend. Moreover, there are a too many "entities collecting and using personal data to make it feasible for people to manage their privacy separately with each entity" [40].

As Cate and Mayer-Schönberger note in the Microsoft Global Privacy Summit Summary Report, "almost everywhere that individuals venture, especially online, they are presented with long and complex privacy notices routinely written by lawyers for lawyers, and then requested to either "consent" or abandon the use of the desired service" [41]. In some cases, organisations try to simplify these policies for the users of their service, but such initiatives make up the minority of terms of use policies. Tene and Polonetsky assert that "it is common knowledge among practitioners in the field that privacy policies serve more as liability disclaimers for businesses than as assurances of privacy for consumers" [42].

However, it is equally important to consider the principle of consent from perspective of companies. At a time where many businesses have to comply with numerous regulations and processes in the name of 'compliance' [43], the obligations for obtaining consent could burden some businesses. Firms have to gather consent amidst enhancing user or customer experiences, which represents a tricky balance to find. For example, requiring consent at every stage may make the user experience much worse. Imagine having to give consent for your profile to be uploaded every time you make a high score in a video game? At the same time, "organizations are expected to explain their data processing activities on increasingly small screens and obtain consent from often-uninterested individuals" [44]. Given these factors, it is somewhat understandable for companies to garner consent for all possible (secondary) uses as otherwise it is not feasible to keep collecting.

Nonetheless, this results in situations where "data processors can perhaps too easily point to the formality of notice and consent and thereby abrogate much of their responsibility" [45]. The totality of the situation shows the odds stacked against the individual. It could be even argued that this is one manifestation of the informational and power asymmetry that exists between individuals and organisations [46], because users may unwittingly agree to unfair, unclear or even unknown terms and conditions and data practices. Not only are individuals greatly misinformed about data collected about them, but the vast majority of people do not even read these Terms and Conditions or End User license agreements [47]. Solove also argues that "people often lack enough expertise to adequately assess the consequences of agreeing to certain present uses or disclosures of their data" [48].

While the organisational practice of providing extensive and complicated terms of use policies is not illegal, the fact that by one estimation, it may take you would have to take 76 working days to review the privacy policies you have agreed to online [49], or by another, that in the USA the opportunity cost society incurs in reading privacy policies is $781 billion [50], should not go unnoticed. I do think it is unfair for the law to put users into such situations, where they are "forced to make overly complex decisions based on limited information" [51]. There have been laudable attempts by some government organisations like Canada's Office of the Privacy Commissioner and USA's Federal Trade Commission to provide guidance to firms to make their privacy policies more accessible [52]. However, these are hard to enforce. Therefore, it can be assumed that when users have neither the expertise nor the rigour to review privacy policies effectively, the consent they provide would naturally be far from informed.

Secondary Use, Aggregation and Superficial Consent

What amplifies this informational asymmetry is the potential for the aggregation of individual's data and subsequent secondary use of that data collected. "Even if people made rational decisions about sharing individual pieces of data in isolation, they greatly struggle to factor in how their data might be aggregated in the future" [53].

This has to do with the prevalence of big data analytics that characterizes our modern epoch, and has major implications for the nature and meaningfulness of the consent users provide. By definition, "big data analysis seeks surprising correlations" [54] and some of its most insightful results are counterintuitive and nearly impossible to conceive at the point of primary data collection. One noteworthy example comes from the USA, with the predictive analytics of Walmart. By studying purchasing patterns of its loyalty card holders [55], the company ascertained that prior to a hurricane the most popular items that people tend to buy are actually Pop Tarts (a pre-baked toaster pastry) and Beer [56]. These correlations are highly counterintuitive and far from what people expect to be necessities before a hurricane. These insights led to Walmart stores being stocked with the most relevant products at the time of need. This is one example of how data might be repurposed and aggregated for a novel purpose, but nonetheless the question about the nature of consent obtained by Walmart for the collection and analysis of the shopping habits of its loyalty card holders stands.

One reason secondary uses make consent less meaningful has been articulated by De Zwart et al, who observe that "the idea of consent becomes unworkable in an environment where it is not known, even by the people collecting and selling data, what will happen to the data" [57]. Taken together with Solove's aggregation effect, two points become apparent:

1. Data we consent to be collected about us may be aggregated with other data we may have revealed in the past. While separately they may be innocuous, there is a risk of future aggregation to create new information which one may find overly intrusive and not consent to. However, current data protection regimes make it hard for one to provide such consent, because there is no way for the user to know how his past and present data may be aggregated in the future.
2. Data we consent to be collected for one specific purpose may be used in a myriad of other ways. The user has virtually no way to know how their data might be repurposed because often time neither do the collectors of that data [58].

Therefore, regulators reliance on principles of purpose limitation and the mechanism of consent for robust data protection seems suboptimal at the very least, as big data practices of aggregation, repurposing and secondary uses become commonplace.

Other Problems with the Mechanism of Consent in the Context of Big Data

On one end of the spectrum are situations where organisations garner consent for future secondary uses at the time of data collection. As discussed earlier, this is currently the common practice for organisations and the likelihood of users providing informed consent is low.

However, equally valid is considering the situations on the other end of the spectrum, where obtaining user consent for secondary use becomes too expensive and cumbersome [59]. As a result, potentially socially valuable secondary use of data for research and innovation or simply "the practice of informed and reflective citizenship" [60] may not take place. While potential social research may be hindered by the consent requirement, the reality that one cannot give meaningful consent to an unknown secondary uses of data is more pressing. Essentially, not knowing what you are consenting to scarcely provides the individual with any semblance of strong privacy protections and so the consent that individuals provide is superficial at best.

Many scholars also point to the binary nature of consent as it stands today [61]. Solove describes consent in data protection as nuanced [62] while Cate and Mayer-Schönberger go further to assert that "binary choice is not what the privacy architects envisioned four decades ago when they imagined empowered individuals making informed decisions about the processing of their personal data". This dichotomous nature of consent further reduces its usefulness in data protection regimes.

Whether data collection is opted into or opted out of also has a bearing on the nature of the consent obtained. Many argue that regulations with options to opt out are not effective as "opt-out consent might be the product of mere inertia or lack of awareness of the option to opt out" [63]. This is in line with initiatives around the world to make gathering consent more explicit by having options to opt in instead of opt out. Noted articulations of the impetus to embrace opt in regimes include ex FTC chairman Jon Leibowitz as early as 2007 [64], as well as being actively considered by the EU in the reform of their data protection laws [65].

However, as Solove rightly points out, opt in consent is problematic as well [66]. There are a few reasons for this: first, that many data collectors have the "sophistication and motivation to find ways to generate high opt-in rates" [67] by "conditioning products, services, or access on opting in" [68]. In essence, they leave individuals no choice but to opt into data collection because using their particular product or service is dependant or 'conditional' on explicit consent. A pertinent example of this is the end-user license agreement to Apple's iTunes Store [69]. Solove rightly notes that "if people want to download apps from the store, they have no choice but to agree. This requirement is akin to an opt-in system — affirmative consent is being sought. But hardly any bargaining or choosing occurs in this process" [70]. Second, as stated earlier, obtaining consent runs the risk of impeding potential innovation or research because it is too cumbersome or expensive to obtain [71].

Third, as Tene and Polonetsky argue, "collective action problems threaten to generate a suboptimal equilibrium where individuals fail to opt into societally beneficial data processing in the hope of free-riding on others' good will" [72]. A useful example to illustrate this comes from another context where obtaining consent is the difference between life and death: organ donation. The gulf in consenting donors between countries with an opt in regime for organ donation and countries with an opt out regime is staggering. Even countries that are culturally similar, such as Austria and Germany, exhibit vast differences in donation rates – Austria at 99% compared to just 12% in Germany [73]. This suggests that in terms of obtaining consent (especially for socially valuable actions), opt in methods may be limiting, because people may have an aversion to anything being presumed about their choices, even if costs of opting out are low [74].

What the above section demonstrates is how consent may be somewhat limited as a tool for data protection regimes, especially in a big data context. That said, consent is not in itself a useless or outdated concept. The problems raised above articulate the problems that relying on consent extensively pose in a big data context. Consent should still remain a part of data protection regimes. However, there are both better ways to obtain consent (for organisations that collect data) as well as other areas to focus regulatory attention on aside from the time of data collection.

What can Organisations do Better to Obtain More Meaningful Consent

Organisations that collect data could alter the way the obtain user consent. Most people can attest to having checked a box that was lying surreptitiously next to the words 'I agree', thereby agreeing to the Terms and Conditions or End-user License Agreement for a particular service or product. This is in line with the need for both parties to assent to the terms of a contract as part of making valid a contract [75]. Some of the more common types of online agreements that users enter into are Clickwrap and Browsewrap agreements. A Clickwrap agreement is "formed entirely in an online environment such as the Internet, which sets forth the rights and obligations between parties" [76]. They "require a user to click "I agree" or "I accept" before the software can be downloaded or installed" [77]. On the other hand, Browsewrap agreements "try to characterize your simple use of their website as your 'agreement' to a set of terms and conditions buried somewhere on the site" [78].

Because Browsewrap agreements do not "require a user to engage in any affirmative conduct" [79], the kind of consent that these types of agreements obtain is highly superficial. In fact, many argue that such agreements are slightly unscrupulous because users are seldom aware that such agreements exist [80], often hidden in small print [81] or below the download button [82] for example. And the courts have begun to consider such terms and practices unfair, which "hold website users accountable for terms and conditions of which a reasonable Internet user would not be aware just by using the site" [83]. For example, In *re Zappos.com Inc., Customer Data Security Breach Litigation*, the court said of their Terms of Use (which is in a browsewrap agreement):

"The Terms of Use is inconspicuous, buried in the middle to bottom of every Zappos.com webpage among many other links, and the website never directs a user to the Terms of Use. No reasonable user would have reason to click on the Terms of Use" [84]

Clearly, courts recognise the potential for consent or assent to be obtained in a hardly transparent or hands on manner. Organisations that collect data should be aware of this and consider other options for obtaining consent.

A few commentators have suggested that organisations switch to using Clickwrap or clickthrough agreements to obtain consent. Undergirding this argument is the fact that courts have on numerous occasions, upheld the validity of a Clickwrap agreement. Such cases include *Groff v. America Online, Inc* [85] and *Hotmail Corporation v. Van Money Pie, Inc* [86]. These cases built upon the precedent-setting case of *Pro CD v. Zeidenberg*, in which the court ruled that "Shrinkwrap licenses are enforceable unless their terms are objectionable on grounds applicable to contracts in general" [87]. Shrinkwrap licenses, which refer to end user license agreements printed on the shrinkwrap of a software product which a user will definitely notice and have the opportunity to read before opening and using the product, and the rules that govern them, have seen application to clickthrough agreements. As Bayley rightly noted, the validity of clickthrough agreements is dependent on "reasonable notice and opportunity to review—whether the placement of the terms and click-button afforded the user a reasonable opportunity to find and read the terms without much effort" [88].

From the perspective of companies and other organisations which attempt to garner consent from users to collect and process their data, utilizing Clickwrap agreements might be one useful solution to consider in obtaining more meaningful and informed consent. In fact Bayley contends that clear Clickwrap agreements are "the "best practice" mechanism for creating a contractual relationship between an online service and a user" [89]. He suggests the following mechanism for acquiring clear and informed consent via contractual agreement [90]:

1. Conspicuously present the TOS to the user prior to any payment (or other commitment by the user) or installation of software (or other changes to a user's machine or browser, like cookies, plug-ins, etc.)
2. Allow the user to easily read and navigate all of the terms (i.e. be in a normal, readable typeface with no scroll box)

3. Provide an opportunity to print, and/or save a copy of, the terms
4. Offer the user the option to decline as prominently and by the same method as the option to agree
5. Ensure the TOS is easy to locate online after the user agrees.

These principles make a lot of sense for organisations, as it requires relatively minor procedural changes instead of more transformational efforts to alter the way the validate their data processing processes entirely.

Herzfield adds two further suggestions to this list. First, organisations should not allow any use of their product or service until "express and active manifestation of assent" [91]. Also, they should institute processes where users re-iterate their consent and assent to the terms of use [92]. He goes further to propose a baseline that organisations should follow: "companies should always provide at least inquiry notice of all terms, and require counterparties to manifest assent, through action or inaction, in a manner that reasonable people would clearly understand to be assent" [93].

While obtaining informed and meaningful consent is neither fool proof nor a process which has widely accepted clear steps, what is clear is that current efforts by organisations may be insufficient. As Cate and Mayer-Schönberger note, "data processors can perhaps too easily point to the formality of notice and consent and thereby abrogate much of their responsibility" [94]. One thing they can do to both ensure more meaningful and informed consent (from the perspective of the users) and preventing potential legal action for unscrupulous or unfair terms is to change the way they obtain consent from opt out to opt in.

Conclusion – How Should Regulation Change

In conclusion, the current emphasis and extensive use of consent in data protection seems to be limited in effectively protecting against illegitimate processing of data in a big data context. More people are starting to use online services extensively. This is coupled by the fact that organisations are realizing the value of collecting and analysing user data to carry out data-driven analytics for insights that can improve the efficacy of the product. Clearly, data protection has never been more crucial.

However not only does emphasising consent seem less relevant, because the consent organisations obtain is seldom informed, but it may even jeopardise the intentions of data protection. Commentators are quick to point out how nimble firms are at acquiring consent in newer ways that may comply with laws but still allow them to maintain their advantageous position of asymmetric power. Kuner, Cate, Millard and Svantesson, all eminent scholars in the field of Big data, asked the prescient question: "Is there a proper role for individual consent?" [95]They believe consent still has a role, but that finding this role in the Big data context is challenging [96]. However, there is surprising consensus on the approach that should be taken as data protection regimes shift away from consent.

In fact, the alternative is staring at us in the face: data protection regimes have to look elsewhere, to other points along the data analysis process for aspects to regulate and ensure legitimate and fair processing of data. One compelling idea which had broad-based support during the aforementioned Microsoft Privacy Summit was that "new approaches must shift responsibility away from data subjects toward data users and toward a focus on accountability for responsible data stewardship" [97], ie creating regulations to guide data processing instead of the data collection. De Zwart et al. suggest that regulation must instead "focus on the processes involved in establishing algorithms and the use of the resulting conclusions" [98].

This might involve regulations relating to requiring data collectors to publish the queries they run on the data. This would be a solution that balances maintaining the 'trade secret' of the firm, who has creatively designed an algorithm, with ensuring fairness and legitimacy in data processing. One manifestation of this approach is in conceptualising procedural data

due process which "would regulate the fairness of Big Data's analytical processes with regard to how they use personal data (or metadata derived from or associated with personal data) in any adjudicative process, including processes whereby Big Data is being used to determine attributes or categories for an individual" [99]. While there is debate regarding the usefulness of a data due process, the idea of data due process is just part of the consortium of ideas surrounding alternatives to consent in data protection. The main point is that "greater transparency should be required if there are fewer opportunities for consent or if personal data can be lawfully collected without consent" [100].

It is also worth considering exactly what a single use of group or individual's data is, and what types of uses or processes require a "greater form of authorization" [101]. Certain data processes could require special affirmative consent to be procured, which is not applicable for other less intimate matters. Canada's Office of the Privacy Commissioner released a privacy toolkit for organisations, in which they provide some exceptions to the consent principle, one of which is if data collection "is clearly in the individual's interests and consent is not available in a timely way" [102]. Some therefore suggest that "if notice and consent are reserved for more appropriate uses, individuals might pay more attention when this mechanism is used" [103].

Another option for regulators is to consider the development and implementation of a sticky privacy policies regime. This refers to "machine-readable policies [that] can stick to data to define allowed usage and obligations as it travels across multiple parties, enabling users to improve control over their personal information" [104]. Sticky privacy policies seem to alleviate the risk of repurposed, unanticipated uses of data because users who consent to giving out their data will be consenting to how it is used thereafter. However, the counter to sticky policies is that it places even greater obligations on users to decide how they would like their data used, not just at one point but for the long term. To expect organisations to state their purposes for future use of individuals data or that individuals are to give informed consent to such uses seems farfetched from both perspectives.

Still another solution draws from the noted scholar Helen Nissenbaum's work on privacy. She argues that "the benchmark of privacy is contextual integrity" [105]. "Contextual integrity ties adequate protection for privacy to norms of specific contexts, demanding that information gathering and dissemination be appropriate to that context and obey the governing norms of distribution within it" [106]. According to this line of thinking, legislators should instead focus their attention on what constitutes appropriateness in certain contexts, although this could be a challenging task as contexts merge and understandings of appropriateness change according to the circumstances of a context. .

While there is little consensus regarding the numerous ways to focus regulatory attention on data processing and the uses of data collected, there is more support for a shift away from consent, as exemplified by the Microsoft privacy Summit:

"There was broad general agreement that privacy frameworks that rely heavily on individual notice and consent are neither sustainable in the face of dramatic increases in the volume and velocity of information flows nor desirable because of the burden they place on individuals to understand the issues, make choices, and then engage in oversight and enforcement" [107]. I think Cate and Mayer- Schönberger make for the most valid conclusion to this article, as well as to summarise the debate I have presented. They say that "in short, ensuring individual control over personal data is not only an increasingly unattainable objective of data protection, but in many settings it is an undesirable one as well" [108]. We might very well be throwing the entire data protection regimes under the bus.

ENDNOTES

[1] Gordon Rayner and Bill Gardner, "Men Must Prove a Woman Said 'Yes' under Tough New Rape Rules - Telegraph," *The Telegraph*, January 28, 2015, sec. Law and Order, http://www.telegraph.co.uk/news/uknews/law-and-order/11375667/Men-must-prove-a-woman-said-Yes-under-tough-new-rape-rules.html.

[2] Legal Information Institute, "Implied Consent," accessed August 25, 2015, https://www.law.cornell.edu/wex/implied_consent.

[3] European Parliament, Council of the European Union, *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 1995*, http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:31995L0046.

[4] See supra note 3.

[5] European Commission, "Stronger Data Protection Rules for Europe," *European Commission Press Release Database*, June 15, 2015, http://europa.eu/rapid/press-release_MEMO-15-5170_en.htm.

[6] Council of the European Union, "Data Protection: Council Agrees on a General Approach," June 15, 2015, http://www.consilium.europa.eu/en/press/press-releases/2015/06/15-jha-data-protection/.

[7] See supra note 6.

[8] Abraham L. Newman, *Protectors of Privacy: Regulating Personal Data in the Global Economy* (Ithaca, NY: Cornell University Press, 2008).

[9] See supra note 8, at 24.

[10] Ibid.

[11] 15 U.S.C. §1681.

[12] 5 U.S.C. § 552a.

[13] 18 U.S.C. § 2510-22.

[14] Federal Trade Commission, "Privacy Online: A Report to Congress," June 1998, https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf: 40.

[15] See supra note 14, at 8.

[16] Organisation for Economic Cooperation and Development, "2013 OECD Privacy Guidelines," 2013, http://www.oecd.org/internet/ieconomy/privacy-guidelines.htm.

[17] Canadian Standards Association, "Canadian Standards Association Model Code," March 1996, https://www.cippguide.org/2010/06/29/csa-model-code/.

[18] Mary Chlopecki, "The Property Rights Origins of Privacy Rights | Foundation for Economic Education," August 1, 1992, http://fee.org/freeman/the-property-rights-origins-of-privacy-rights.

[19] See *Pope v. Curl* (1741), available here.

[20] See *Prince Albert v. Strange* (1849), available here.

[21] Samuel D. Warren and Louis D. Brandeis, "The Right to Privacy," *Harvard Law Review* 4, no. 5 (December 15, 1890): 193–220, doi:10.2307/1321160.

[22] See supra note 18.

[23] Ibid.

[24] See supra note 21.

[25] See for example, Richard Epstein, "Privacy, Property Rights, and Misrepresentations," *Georgia Law Review*, January 1, 1978, 455. And Richard Posner, "The Right of Privacy," *Sibley Lecture Series*, April 1, 1978, http://digitalcommons.law.uga.edu/lectures_pre_arch_lectures_sibley/22.

[26] See supra note 21, at 215.

[27] See supra note 21, at 218.

[28] Ibid.

[29] Adrienne W. Fawcett, "Q: Who Said: 'A Man's Home Is His Castle'?," *Chicago Tribune*, September 14, 1997, http://articles.chicagotribune.com/1997-09-14/news/9709140446_1_castle-home-sir-edward-coke.

[30] Brendan Purves, "Castle Doctrine from State to State," *South Source*, July 15, 2011, http://source.southuniversity.edu/castle-doctrine-from-state-to-state-46514.aspx.

[31] "Volenti Non Fit Injuria," *E-Lawresources*, accessed August 25, 2015, http://e-lawresources.co.uk/Volenti-non-fit-injuria.php.

[32] Bryce Clayton Newell, "Local Law Enforcement Jumps on the Big Data Bandwagon: Automated License Plate Recognition Systems, Information Privacy, and Access to Government Information," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, October 16, 2013), http://papers.ssrn.com/abstract=2341182.

[33] Alan Westin, *Privacy and Freedom* (Ig Publishing, 2015).

[34] Helen Nissenbaum, "Privacy as Contextual Integrity," *Washington Law Review* 79 (2004): 119.

[35] Ruth Gavison, "Privacy and the Limits of Law," *The Yale Law Journal* 89, no. 3 (January 1, 1980): 421–71, doi:10.2307/795891: 423.

[36] Ibid.

[37] Daniel J. Solove, "Privacy Self-Management and the Consent Dilemma," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 4, 2012), http://papers.ssrn.com/abstract=2171018: 1888.

[38] Ibid, at 1889.

[39] Omer Tene and Jules Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, September 20, 2012),http://papers.ssrn.com/abstract=2149364: 261.

[40] See supra note 37, at 1881.

[41] Fred H. Cate and Viktor Mayer-Schönberger, "Notice and Consent in a World of Big Data - Microsoft Global Privacy Summit Summary Report and Outcomes," Microsoft Global Privacy Summit, November 9, 2012,http://www.microsoft.com/en-us/download/details.aspx?id=35596: 3.

[42] See supra note 39.

[43] See for example, US Securities and Exchange Commission, "Corporation Finance Small Business Compliance Guides," accessed August 26, 2015, https://www.sec.gov/info/smallbus/secg.shtml and Australian Securities & Investments Commission, "Compliance for Small Business," accessed August 26, 2015, http://asic.gov.au/for-business/your-business/small-business/compliance-for-small-business/.

[44] See supra note 39.

[45] See supra note 41.

[46] See supra note 8, at 24.

[47] See for example, James Daley, "Don't Waste Time Reading Terms and Conditions," *The Telegraph*, September 3, 2014, and Robert Glancy, "Will You Read This Article about Terms and Conditions? You Really Should Do," *The Guardian*, April 24, 2014, sec. Comment is free, http://www.theguardian.com/commentisfree/2014/apr/24/terms-and-conditions-online-small-print-information.

[48] See supra note 37, at 1886.

[49] Alex Hudson, "Is Small Print in Online Contracts Enforceable?," *BBC News*, accessed August 26, 2015, http://www.bbc.com/news/technology-22772321.

[50] Aleecia M. McDonald and Lorrie Faith Cranor, "Cost of Reading Privacy Policies, The," *I/S: A Journal of Law and Policy for the Information Society* 4 (2009 2008): 541

[51] See supra note 41, at 4.

[52] For Canada, see Office of the Privacy Commissioner of Canada, "Fact Sheet: Ten Tips for a Better Online Privacy Policy and Improved Privacy Practice Transparency," October 23, 2013, https://www.priv.gc.ca/resource/fs-fi/02_05_d_56_tips2_e.asp. And Office of the Privacy Commissioner of Canada, "Privacy Toolkit - A Guide for Businesses and Organisations to Canada's Personal Information Protection and Electronic Documents Act," accessed August 26, 2015, https://www.priv.gc.ca/information/pub/guide_org_e.pdf.
For USA, see Federal Trade Commission, "Internet of Things: Privacy & Security in a Connected World," Staff Report (Federal Trade Commission, January 2015), https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf.

[53] See supra note 37, at 1889.

[54] See supra note 39, at 261.

[55] Jakki Geiger, "The Surprising Link Between Hurricanes and Strawberry Pop-Tarts: Brought to You by Clean, Consistent and Connected Data," *The Informatica Blog - Perspectives for the Data Ready Enterprise*, October 3, 2014, http://blogs.informatica.com/2014/03/10/the-surprising-link-between-strawberry-pop-tarts-and-hurricanes-brought-to-you-by-clean-consistent-and-connected-data/#fbid=PElJO4Z_kOu.

[56] Constance L. Hays, "What Wal-Mart Knows About Customers' Habits," *The New York Times*, November 14, 2004, http://www.nytimes.com/2004/11/14/business/yourmoney/what-walmart-knows-about-customers-habits.html.

[57] M. J. de Zwart, S. Humphreys, and B. Van Dissel, "Surveillance, Big Data and Democracy: Lessons for Australia from the US and UK," Http://www.unswlawjournal.unsw.edu.au/issue/volume-37-No-2, 2014,https://digital.library.adelaide.edu.au/dspace/handle/2440/90048: 722.

[58] Ibid.

[59] See supra note 41, at 3.

[60] Julie E. Cohen, "What Privacy Is For," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 5, 2012), http://papers.ssrn.com/abstract=2175406.

[61] See supra note 37, at 1901.

[62] Ibid.

[63] See supra note 37, at 1899.

[64] Jon Leibowitz, "So Private, So Public: Individuals, The Internet & The paradox of behavioural marketing" November 1, 2007, https://www.ftc.gov/sites/default/files/documents/public_statements/so-private-so-public-individuals-internet-paradox-behavioral-marketing/071031ehavior_0.pdf: 6.

[65] See supra note 5.

[66] See supra note 37, at 1898.

[67] Ibid.

[68] Ibid.

[69] Ibid.

[70] Ibid.

[71] See supra note 41, at 3.

[72] See supra note 39, at 261.

[73] Richard H. Thaler, "Making It Easier to Register as an Organ Donor," *The New York Times*, September 26, 2009, http://www.nytimes.com/2009/09/27/business/economy/27view.html.

[74] Ibid.

[75] *The Oxford Introductions to U.S. Law: Contracts*, 1 edition (New York: Oxford University Press, 2010): 67.

[76] Francis M. Buono and Jonathan A. Friedman, "Maximizing the Enforceability of Click-Wrap Agreements," *Journal of Technology Law & Policy* 4, no. 3 (1999), http://jtlp.org/vol4/issue3/friedman.html.

[77] North Carolina State University, "Clickwraps," *Software @ NC State Information Technology*, accessed August 26, 2015, http://software.ncsu.edu/clickwraps.

[78] Ed Bayley, "The Clicks That Bind: Ways Users 'Agree' to Online Terms of Service," *Electronic Frontier Foundation*, November 16, 2009, https://www.eff.org/wp/clicks-bind-ways-users-agree-online-terms-service.

[79] Ibid, at 2.

[80] Ibid.

[81] See *Nguyen v. Barnes & Noble Inc.*, (9th Cir. 2014), available at http://cdn.ca9.uscourts.gov/datastore/opinions/2014/08/18/12-56628.pdf

[82] See Specht v. Netscape Communications Corp.,(2d Cir. 2002), available at http://cyber.law.harvard.edu/stjohns/Specht_v_Netscape.pdf

[83] See supra note 78, at 2.

[84] See In Re: Zappos.com, Inc., Customer Data Security Breach Litigation, No. 3:2012cv00325: pg 8 line 23-26, available at http://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=1152&context=historical

[85] See Groff v. America Online, Inc., 1998, available at http://www.internetlibrary.com/cases/lib_case20.cfm

[86] Hotmail Corp. v. Van$ Money Pie, Inc., 1998, available at https://cyber.harvard.edu/property00/alternatives/hotmail.html

[87] ProCD Inc. v. Zeidenberg, (7th. Cir. 1996), available at https://www.law.cornell.edu/copyright/cases/86_F3d_1447.htm

[88] See supra note 78, at 1.

[89] See supra note 78, at 2.

[90] Ibid.

[91] Oliver Herzfeld, "Are Website Terms Of Use Enforceable?," *Forbes*, January 22, 2013, http://www.forbes.com/sites/oliverherzfeld/2013/01/22/are-website-terms-of-use-enforceable/.

[92] Ibid.

[93] Ibid.

[94] See supra note 41, at 3.

[95] Christopher Kuner et al., "The Challenge of 'big Data' for Data Protection," *International Data Privacy Law* 2, no. 2 (May 1, 2012): 47–49, doi:10.1093/idpl/ips003: 49.

[96] Ibid.

[97] See supra note 41, at 5.

[98] See supra note 57, at 723.

[99] Kate Crawford and Jason Schultz, "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, October 1, 2013), http://papers.ssrn.com/abstract=2325784: 109.

[100] See supra note 41, at 13.

[101] See supra note 41, at 5.

[102] See supra note 52, Privacy Toolkit, at 14.

[103] See supra note 41, at 6.

[104] Siani Pearson and Marco Casassa Mont, "Sticky Policies: An Approach for Managing Privacy across Multiple Parties," *Computer*, 2011.

[105] See supra note 34, at 138.

[106] See supra note 34, at 118.

[107] See supra note 41, at 5.

[108] See supra note 41, at 4.

A Critique of Consent in Information Privacy

AMBER SINHA AND SCOTT MASON

Notice and Consent as Cornerstone of Privacy Law

The privacy notice, which is the primary subject of this article, conveys all pertinent information, including risks and benefits to the participant, and in the possession of such knowledge, they can make an informed choice about whether to participate or not.

Most modern laws and data privacy principles seek to focus on individual control. In this context, the definition by the late Alan Westin, former Professor of Public Law & Government Emeritus, Columbia University, which characterises privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to other," [1] is most apt. The idea of privacy as control is what finds articulation in data protection policies across jurisdictions beginning from the Fair Information Practice Principles (FIPP) from the United States [2]. Paul Schwarz, the Jefferson E. Peyser Professor at UC Berkeley School of Law and a Director of the Berkeley Center for Law and Technology, called the FIPP the building blocks of modern information privacy law [3]. These principles trace their history to a report called 'Records, Computers and Rights of Citizens' [4] prepared by an Advisory Committee appointed by the US Department of Health, Education and Welfare in 1973 in response to the increasing automation in data systems containing information about individuals. The Committee's mandate was to "explore the impact of computers on record keeping about individuals and, in addition, to inquire into, and make recommendations regarding, the use of the Social Security number" [5]. The most important legacy of this report was the articulation of five principles which would not only play a significant role in the privacy laws in US but also inform data protection law in most privacy regimes internationally [6] like the OECD Privacy Guidelines, the EU Data Protection Principles, the FTC Privacy Principles, APEC Framework or the nine National Privacy Principles articulated by the Justice A P Shah Committee Report which are reflected in the Privacy Bill, 2014 in India. Fred Cate, the C. Ben Dutton Professor of Law at the Indiana University Maurer School of Law, effectively summarises the import of all of these privacy regimes as follows:

"All of these data protection instruments reflect the same approach: tell individuals what data you wish to collect or use, give them a choice, grant them access, secure those data with appropriate technologies and procedures, and be subject to third-party enforcement if you fail to comply with these requirements or individuals' expressed preferences" [7]

This makes the individual empowered and allows them to weigh their own interests in exercising their consent. The allure of this paradigm is that in one elegant stroke, it seeks to "ensure that consent is informed and free and thereby also to implement an acceptable tradeoff between privacy and competing concerns" [8]. This system was originally intended to be only one of the multiple ways in data processing would be governed, along with other substantive principles such as data quality, however, it soon became the dominant and often the only mechanism [9]. In recent years however, the emergence of Big Data and the nascent development of the Internet of Things has led many commentators to begin questioning the workability of consent as a principle of privacy [10]. In this article we will look closely at the some of issues with the concept of informed consent, and how these notions have become more acute in recent years. Following an analysis of these issues, we will conclude by arguing that today consent, as the cornerstone of privacy law, may in fact be thought of as counterproductive and that a rethinking of a principle based approach to privacy may be necessary.

Problems with Consent

To a certain extent, there are some cognitive problems that have always existed with the issue of informed consent such as long and difficult to understand privacy notices [11], although, in recent past with these problems have become much more aggravated. Fred Cate points out that FIPPs at their inception were broad principles which included both substantive and procedural aspects. However, as they were translated into national laws, the emphasis remained on the procedural aspect of notice and consent. From the idea of individual or societal welfare as the goals of privacy, the focus had shifted to individual control [12]. With data collection occurring with every use of online services, and complex data sets being created, it is humanly impossible to exercise rational decision-making about the choice to allow someone to use our personal data. The thrust of Big Data technologies is that the value of data resides not in its primary purposes but in its numerous secondary purposes where data is re-used many times over [13]. In that sense, the very idea of Big Data conflicts with the data minimization principle [14]. The idea is to retain as much data as possible for secondary uses. Since, these secondary uses are, by their nature, unanticipated, its runs counter to the the very idea of the purpose limitation principle [15]. The notice and consent requirement has simply led to a proliferation of long and complex privacy notices which are seldom read and even more rarely understood. We will articulate some issues with privacy notices which have always existed, and have only become more exacerbated in the context of Big Data and the Internet of Things.

1. Failure to Read/ Access Privacy Notices

The notice and consent principle relies on the ability of the individual to make an informed choice after reading the privacy notice. The purpose of a privacy notice is to act as a public announcement of the internal practices on collection, processing, retention and sharing of information and make the user aware of the same [16]. However, in order to do so the individual must first be able to access the privacy notices in an intelligible format and read them. Privacy notices come in various forms, ranging from documents posted as privacy policies on a website, to click through notices in a mobile app, to signs posted in public spaces informing about the presence of CCTV cameras [17].

In order for the principle of notice and consent to work, the privacy notices need to be made available in a language understood by the user. As per estimates, about 840 million people (11% of the world population) can speak or understand English. However, most privacy notices online are not available in the local language in different regions [18]. Further, with the ubiquity of smartphones and advent of Internet of Things, constrained interfaces on mobile screens and wearables make the privacy notices extremely difficult to read. It must be remembered that privacy notices often run into several pages, and smaller screens effectively ensure that most users do not read through them. Further, connected wearable devices often have “little or no interfaces that readily permit choices” [19]. As more and more devices are connected, this problem will only get more pronounced. Imagine in a world where refrigerators act as the intermediary disclosing information to your doctor or supermarket, at what point does the data subject step in and exercise consent [20].

Another aspect that needs to be understood is that unlike earlier when data collectors were far and few in between, the user could theoretically make a rational choice taking into account the purpose of data collection. However, in the world of Big Data, consent often needs to be provided while the user is trying to access services. In that context click through privacy notices such as those required to access online application, are treated simply as an impediment that must be crossed in order to get access to services. The fact that the consent need to be given in real time almost always results in disregarding what the privacy notices say [21].

Finally, some scholars have argued that while individual control over data may be appealing in theory, it merely gives an illusion of enhanced privacy but not the reality of meaningful choice [22]. Research demonstrates that the presence of the term ‘privacy policy’

leads peopleto the false assumption that if a company has a privacy policy in place, it automatically means presence of substantive and responsible limits on how data is handled [23]. Joseph Turow, the Robert Lewis Shayon Professor of Communication at the Annenberg School for Communication, and his team for example has demonstrated how "[w]hen consumers see the term 'privacy policy,' they believe that their personal information will be protected in specific ways; in particular, they assume that a website that advertises a privacy policy will not share their personal information" [24]. In reality, however, privacy policies are more likely to serve as liability disclaimers for companies than any kind of guarantee of privacy for consumers. Most people tend to ignore privacy policies [25]. Cass Sunstein states that our cognitive capacity to make choices and take decisions is limited. When faced with an overwhelming number of choices to make, most of us do not read privacy notices and resort to default options [26]. The requirement to make choices, sometimes several times in a day, imposes significant burden on the consumers as well the business seeking such consent [27].

2. Failure to Understand Privacy Notices

FTC chairperson Edith Ramirez stated: "In my mind, the question is not whether consumers should be given a say over unexpected uses of their data; rather, the question is how to provide simplified notice and choice" [28]. Privacy notices often come in the form of long legal documents much to the detriment of the readers' ability to understand them. These policies are "long, complicated, full of jargon and change frequently" [29]. Kent walker list five problems that privacy notices typically suffer from - a) overkill - long and repetitive text in small print, b) irrelevance - describing situations of little concern to most consumers, c) opacity - broad terms the reflect the truth that is impossible to track and control all the information collected and stored, d) non-comparability - simplification required to achieve comparability will lead to compromising accuracy, and e) inflexibility - failure to keep pace with new business models [30]. Erik Sherman did a review of twenty three corporate privacy notices and mapped them against three indices which give approximate level of education necessary to understand text on a first read. His results show that most of policies can only be understood on the first read by people of a grade level of 15 or above [31]. FTC Chairperson Timothy Muris summed up the problem with long privacy notices when he said, "Acres of trees died to produce a blizzard of barely comprehensible privacy notices" [32].

Margaret Jane Radin, the former Henry King Ransom Professor of Law Emerita at the University of Michigan, provides a good definition of free consent. It "involves a knowing understanding of what one is doing in a context in which it is actually possible for or to do otherwise, and an affirmative action in doing something, rather than a merely passive acquiescence in accepting something" [33]. There have been various proposals advocating a more succinct and simpler standard for privacy notices [34], or multi-layered notices [35] or representing the information in the form of a table [36]. However, studies show only an insignificant improvement in the understanding by consumers when privacy policies are represented in graphic formats like tables and labels [37]. It has also been pointed out that it is impossible to convey complex data policies in simple and clear language [38].

3. Failure to Anticipate/Comprehend the Consequences of Consent

Today's infinitely complex and labyrinthine data ecosystem is beyond the comprehension of most ordinary users. Despite a growing willingness to share information online, most have no understanding of what happens to their data once they have uploaded it - Where it goes? Whom it is held by? Under what conditions? For what purpose? Or how might it be used, aggregated, hacked, or leaked in the future? For the most part, the above operations are "invisible, managed at distant centers, from behind the scenes, by unmanned powers" [39].

The perceived opportunities and benefits of Big Data have led to an acceptance of the indiscriminate collection of as much data as possible as well as the retention of that data for unspecified future analysis. For many advocates, such practices are absolutely essential if Big Data is to deliver on its promises.. Experts have argued that key privacy principles

particularly those of collection limitation, data minimization and purpose limitation should not be applied to Big Data processing [40]. As mentioned above, in the case of Big Data, the value of the data collected comes often not from its primary purpose but from its secondary uses. Deriving value from datasets involves amalgamating diverse datasets and executing speculative and exploratory kinds of analysis in order to discover hidden insights and correlations that might have previously gone unnoticed [41]. As such organizations are today routinely reprocessing data collected from individuals for purposes not directly related to the services they provide to the customer. These secondary uses of data are becoming increasingly valuable sources of revenue for companies as the value of data in and of itself continues to rise [42].

Purpose Limitation

The principle of purpose limitation has served as a key component of data protection for decades. Purposes given for the processing of users' data should be given at the time of collection and consent and should be "specified, explicit and legitimate". In practice however, reasons given typically include phrases such as, 'for marketing purposes' or 'to improve the user experience' that are vague and open to interpretation [43].

Some commentators whilst conceding the fact that purpose limitation in the era of Big Data may not be possible have instead attempted to emphasise the notion of 'compatible use' requirements. In the view of Working Party on the protection of individuals with regard to the processing of person data, for example, use of data for a purpose other than that originally stated at the point of collection should be subject to a case-by-case review of whether not further processing for different purpose is justifiable - i.e., compatible with the original purpose. Such a review may take into account for example, the context in which the data was originally collected, the nature or sensitivity of the data involved, and the existence of relevant safeguards to insure fair processing of the data and prevent undue harm to the data subject [44].

On the other hand, Big Data advocates have argued that an assessment of legitimate interest rather than compatibility with the initial purpose is far better suited to Big Data processing [45]. They argue that today the notion of purpose limitation has become outdated. Whereas previously data was collected largely as a by-product of the purpose for which it was being collected. If for example, we opted to use a service the information we provided was for the most part necessary to enable the provision of that service. Today however, the utility of data is no longer restricted to the primary purpose for which it is collected but can be used to provide all kinds of secondary services and resources, reduce waste, increase efficiency and improve decision-making [46]. These kinds of positive externalities, Big Data advocates insist, are only made possible by the reprocessing of data.

Unfortunately for the notion of consent the nature of these secondary purposes are rarely evident at the time of collection. Instead the true value of the data can often only be revealed when it is amalgamated with other diverse datasets and subjected to various forms of analysis to help reveal hidden and non-obvious correlations and insights [47]. The uncertain and speculative value of data therefore means that it is impossible to provide "specific, explicit, and legitimate" details about how a given data set will be used or how it might be aggregated in future. Without this crucial information data subjects have no basis upon which they can make an informed decision about whether or not to provide consent. Robert Sloan and Richard Warner argue that it is impossible for a privacy notice to contain enough information to enable free consent. They argue that current data collection practices are highly complex and that these practices involve collection of information at one stage for one purpose and then retain, analyze, and distribute it for a variety of other purposes in unpredictable ways [48]. Helen Nissenbaum points to the ever changing nature of data flow and the cognitive challenges it poses. "Even if, for a given moment, a snapshot of the information flows could be grasped, the realm is in constant flux, with new firms entering the picture, new analytics, and new back end contracts forged: in other words, we are dealing with a recursive capacity that is indefinitely extensible" [49].

Scale and Aggregation

Today the quantity of data being generated is expanding at an exponential rate. From smartphones and televisions, trains and airplanes, sensor-equipped buildings and even the infrastructures of our cities, data now streams constantly from almost every sector and function of daily life, 'creating countless new digital puddles, lakes, tributaries and oceans of information' [50]. In 2011 it was estimated that the quantity of data produced globally would surpass 1.8 zettabytes , by 2013 that had grown to 4 zettabytes, and with the nascent development of the Internet of Things gathering pace, these trends are set to continue [51]. Big Data by its very nature requires the collection and processing of very large and very diverse data sets. Unlike other forms scientific research and analysis which utilize various sampling techniques to identify and target the types of data most useful to the research questions, Big Data instead seeks to gather as much data as possible, in order to achieve full resolution of the phenomenon being studied, a task made much easier in recent years as a result of the proliferation of internet enabled devices and the growth of the Internet of Things. This goal of attaining comprehensive coverage exists in tension however with the key privacy principles of collection limitation and data minimization which seek to limit both the quantity and variety of data collected about an individual to the absolute minimum [52].

The dilution of the purpose limitation principle entails that even those who understand privacy notices and are capable of making rational choices about it, cannot conceptualize how their data will be aggregated and possibly used or re-used. Seemingly innocuous bits of data revealed at different stages could be combined to reveal sensitive information about the individual. Daniel Solove, the John Marshall Harlan Research Professor of Law at the George Washington University Law School, in his book, "The Digital Person", calls it the aggregation effect. He argues that the ingenuity of the data mining techniques and the insights and predictions that could be made by it render any cost-benefit analysis that an individual could make ineffectual [53].

4. Failure to Opt-Out

The traditional choice against the collection of personal data that users have had access to, at least in theory, is the option to 'opt-out' of certain services. This draws from the free market theory that individuals exercise their free will when they use services and always have the option of opting out, thus, arguing against regulation but relying on the collective wisdom of the market to weed out harms. The notion that the provision of data should be a matter of personal choice on the part of the individual and that the individual can, if they chose decide to 'opt-out' of data collection, for example by ceasing use of a particular service, is an important component of privacy and data protection frameworks. The proliferation of internet-enabled devices, their integration into the built environment and the real-time nature of data collection and analysis however are beginning to undermine this concept. For many critics of Big Data, the ubiquity of data collection points as well as the compulsory provision of data as a prerequisite for the access and use of many key online services, is making opting-out of data collection not only impractical but in some cases impossible [54].

Whilst sceptics may object that individuals are still free to stop using services that require data. As online connectivity becomes increasingly important to participation in modern life, the choice to withdraw completely is becoming less of a genuine choice [55]. Information flows not only from the individuals it is about but also from what other people say about them. Financial transactions made online or via debit/credit cards can be analysed to derive further information about the individual. If opting-out makes you look anti-social, criminal, or unethical, the claims that we are exercising free will seems murky and leads one to wonder whether we are dealing with coercive technologies.

Another issue with the consent and opt-out paradigm is the binary nature of the choice. This binary nature of consent makes a mockery of the notion that consent can function as an effective tool of personal data management. What it effectively means is that one can either

agree with the long privacy notices, or choose to abandon the desired service. “This binary choice is not what the privacy architects envisioned four decades ago when they imagined empowered individuals making informed decisions about the processing of their personal data. In practice, it certainly is not the optimal mechanism to ensure that either information privacy or the free flow of information is being protected” [56].

Conclusion: ‘Notice and Consent’ is Counter-Productive

There continues to be an unwillingness amongst many privacy advocates to concede that the concept of consent is fundamentally broken, as Simon Davies, a privacy advocate based in London, comments ‘to do so could be seen as giving ground to the data vultures’, and risks further weakening an already dangerously fragile privacy framework [57]. Nevertheless, as we begin to transition into an era of ubiquitous data collection, evidence is becoming stronger that consent is not simply ineffective, but may in some instances might be counterproductive to the goals of privacy and data protection.

As already noted, the notion that privacy agreements produce anything like truly informed consent has long since been discredited; given this fact, one may ask for whose benefit such agreements are created? One may justifiably argue that far from being for the benefit and protection of users, privacy agreement may in fact be fundamentally to the benefit of data brokers, who having gained the consent of users can act with near impunity in their use of the data collected. Thus, an overly narrow focus on the necessity of consent at the point of collection, risks diverting our attention from the arguably more important issue of how our data is stored, analysed and distributed by data brokers following its collection [58].

Furthermore, given the often complicated and cumbersome processes involved in gathering consent from users, some have raised concerns that the mechanisms put in place to garner consent could themselves morph into surveillance mechanisms. Davies, for example cites the case of the EU Cookie Directive, which required websites to gain consent for the collection of cookies. Davies observes how, ‘a proper audit and compliance element in the system could require the processing of even more data than the original unregulated web traffic. Even if it was possible for consumers to use some kind of gateway intermediary to manage the consent requests, the resulting data collection would be overwhelming”. Thus in many instances there exists a fundamental tension between the requirement placed on companies to gather consent and the equally important principle of data minimization [59].

Given the above issues with notice and informed consent in the context of information privacy, and the fact that it is counterproductive to the larger goals of privacy law, it is important to revisit the principle or rights based approach to data protection, and consider a paradigm shift where one moves to a risk based approach that takes into account the actual threats of sharing data rather than relying on what has proved to be an ineffectual system of individual control. We will be dealing with some of these issues in a follow up to this article.

ENDNOTES

[1] Alan Westin, Privacy and Freedom, Atheneum, New York, 2015.

[2] FTC Fair Information Practice Principles (FIPP) available at https://www.it.cornell.edu/policies/infoprivacy/principles.cfm.

[3] Paul M. Schwartz, “Privacy and Democracy in Cyberspace,” 52 Vanderbilt Law Review 1607, 1614 (1999).

[4] US Secretary’s Advisory Committee on Automated Personal Data Systems, Records, Computers and the Rights of Citizens, available at http://www.justice.gov/opcl/docs/rec-com-rights.pdf

[5] https://epic.org/privacy/ppsc1977report/c13.htm .

[6] Marc Rotenberg, “Fair Information Practices and the Architecture of Privacy: What Larry Doesn’t Get,” available at https://journals.law.stanford.edu/sites/default/files/stanford-technology-law-review/online/rotenberg-fair-info-practices.pdf

[7] Fred Cate, The Failure of Information Practice Principles, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1156972

[8] Robert Sloan and Richard Warner, Beyong Notice and Choice: Privacy, Norms and Consent, 2014, available at https://www.suffolk.edu/documents/jhtl_publications/SloanWarner.pdf

[9] Fred Cate, Viktor Schoenberger, Notice and Consent in a world of Big Data, available at http://idpl.oxfordjournals.org/content/3/2/67.abstract

[10] Daniel Solove, Privacy self-management and consent dilemma, 2013 available athttp://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2093&context=faculty_publications

[11] Ben Campbell, Informed consent in developing countries: Myth or Reality, available athttps://www.dartmouth.edu/~ethics/docs/Campbell_informedconsent.pdf ;

[12] Supra Note 7.

[13] Viktor Mayer Schoenberger and Kenneth Cukier, Big Data: A Revolution that will transform how we live, work and think" John Murray, London, 2013 at 153.

[14] The Data Minimization principle requires organizations to limit the collection of personal data to the minimum extent necessary to obtain their legitimate purpose and to delete data no longer required.

[15] Omer Tene and Jules Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," SSRN Scholarly Paper, available at http://papers.ssrn.com/abstract=2149364

[16] Florian Schaub, R. Balebako et al, "A Design Space for effective privacy notices" available at https://www.usenix.org/system/files/conference/soups2015/soups15-paper-schaub.pdf

[17] Daniel Solove, The Digital Person: Technology and Privacy in the Information Age, NYU Press, 2006.

[18] http://www.ethnologue.com/statistics/size

[19] Opening Remarks of FTC Chairperson Edith Ramirez Privacy and the IoT: Navigating Policy Issues International Consumer Electronics Show Las Vegas, Nevada January 6, 2015 available at https://www.ftc.gov/system/files/documents/public_statements/617191/150106cesspeech.pdf

[20] http://www.privacysurgeon.org/blog/incision/why-the-idea-of-consent-for-data-processing-is-becoming-meaningless-and-dangerous/

[21] Supra Note 10.

[22] Supra Note 7.

[23] Chris Jay Hoofnagle & Jennifer King, Research Report: What Californians Understand About Privacy Online, available at http://ssrn.com/abstract=1262130

[24] Joseph Turrow, Michael Hennesy, Nora Draper, The Tradeoff Fallacy, available at https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf

[25] Saul Hansell, "Compressed Data: The Big Yahoo Privacy Storm That Wasn't," New York Times, May 13, 2002 available at http://www.nytimes.com/2002/05/13/business/compressed-data-the-big-yahoo-privacy-storm-that-wasn-t.html?_r=0

[26] Cass Sunstein, Choosing not to choose: Understanding the Value of Choice, Oxford University Press, 2015.

[27] For example, Acxiom, processes more than 50 trillion data transactions a year. http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html?pagewanted=all&_r=0

[28] Opening Remarks of FTC Chairperson Edith Ramirez Privacy and the IoT: Navigating Policy Issues International Consumer Electronics Show Las Vegas, Nevada January 6, 2015 available at https://www.ftc.gov/system/files/documents/public_statements/617191/150106cesspeech.pdf

[29] L. F. Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. Journal on Telecommunications and High Technology Law, 10:273, 2012, available at http://jthtl.org/content/articles/V10I2/JTHTLv10i2_Cranor.PDF

[30] Kent Walker, The Costs of Privacy, 2001 available at https://www.questia.com/library/journal/1G1-84436409/the-costs-of-privacy

[31] Erik Sherman, "Privacy Policies are great - for Phds", CBS News, available at http://www.cbsnews.com/news/privacy-policies-are-great-for-phds/

[32] Timothy J. Muris, Protecting Consumers' Privacy: 2002 and Beyond, available at http://www.ftc.gov/speeches/muris/privisp1002.htm

[33] Margaret Jane Radin, Humans, Computers, and Binding Commitment, 1999 available at http://www.repository.law.indiana.edu/ilj/vol75/iss4/1/

[34] Annie I. Anton et al., Financial Privacy Policies and the Need for Standardization, 2004 available athttps://ssl.lu.usi.ch/entityws/Allegati/pdf_pub1430.pdf; Florian Schaub, R. Balebako et al, "A Design Space for effective privacy notices" available at https://www.usenix.org/system/files/conference/soups2015/soups15-paper-schaub.pdf

[35] The Center for Information Policy Leadership, Hunton & Williams LLP, "Ten Steps To Develop A Multi-Layered Privacy Notice" available at https://www.informationpolicycentre.com/files/Uploads/Documents/Centre/Ten_Steps_whitepaper.pdf

[36] Allen Levy and Manoj Hastak, Consumer Comprehension of Financial Privacy Notices, Interagency Notice Project, available at https://www.sec.gov/comments/s7-09-07/s70907-21-levy.pdf

[37] Patrick Gage Kelly et al., Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach available at https://www.ftc.gov/sites/default/files/documents/public_comments/privacy-roundtables-comment-project-no.p095416-544506-00037/544506-00037.pdf

[38] Howard Latin, "Good" Warnings, Bad Products, and Cognitive Limitations, 41 UCLA Law Review available at https://litigation-essentials.lexisnexis.com/webcd app?action=DocumentDisplay&crawlid=1&srctype=smi&srcid=3B15&doctype=cite&docid=41+UCLA+L.+Rev.+1193&key=1c15e064a97759f3f03fb51db62a79a5

[39] Jonathan Obar, Big Data and the Phantom Public: Walter Lippmann and the fallacy of data privacy self management, Big Data and Society, 2015, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2239188

[40] Viktor Mayer Schoenberger and Kenneth Cukier, Big Data: A Revolution that will transform how we live, work and think" John Murray, London, 2013.

[41] Supra Note 15.

[42] Supra Note 40.

[43] Article 29 Working Party, (2013) Opinion 03/2013 on Purpose Limitation, Article 29, available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

[44] Ibid.

[45] It remains unclear however whose interest would be accounted, existing EU legislation would allow commercial/data broker/third party interests to trump those of the user, effectively allowing re-processing of personal data irrespective of whether that processing would be in the interest of the user.

[46] Supra Note 40.

[47] Supra Note 10.

[48] Robert Sloan and Richard Warner, Beyong Notice and Choice: Privacy, Norms and Consent, 2014, available at https://www.suffolk.edu/documents/jhtl_publications/SloanWarner.pdf

[49] Helen Nissenbaum, A Contextual Approach to Privacy Online, available at http://www.amacad.org/publications/daedalus/11_fall_nissenbaum.pdf

[50] D Bollier, The Promise and Peril of Big Data. The Aspen Institute, 2010, available at: http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf

[51] Meeker, M. & Yu, L. Internet Trends, Kleiner Perkins Caulfield Byers, (2013), http://www.slideshare.net/kleinerperkins/kpcb-internet-trends-2013 .

[52] Supra Note 40.

[53] Supra Note 17.

[54] Janet Vertasi, My Experiment Opting Out of Big Data Made Me Look Like a Criminal, 2014, available at http://time.com/83200/privacy-internet-big-data-opt-out/

[55] Ibid.

[56] http://www.techpolicy.com/NoticeConsent-inWorldBigData.aspx

[57] Simon Davies, Why the idea of consent for data processing is becoming meaningless and dangerous, available at http://www.privacysurgeon.org/blog/incision/why-the-idea-of-consent-for-data-processing-is-becoming-meaningless-and-dangerous/

[58] Supra Note 10.

[59] Simon Davies, Why the idea of consent for data processing is becoming meaningless and dangerous, available at http://www.privacysurgeon.org/blog/incision/why-the-idea-of-consent-for-data-processing-is-becoming-meaningless-and-dangerous/

A Case for Greater Privacy Paternalism?

AMBER SINHA

Background

The current data privacy protection framework across most jurisdictions is built around a rights based approach which entrusts the individual with having the wherewithal to make informed decisions about her interests and well-being [1]. In his book, *The Phantom Public*, published in 1925, Walter Lippmann argues that the rights based approach is based on the idea of a sovereign and omnicompetent citizens, who can direct public affairs, however, this idea is a mere phantom or an abstraction [2]. Jonathan Obar, Assistant Professor of Communication and Digital Media Studies in the Faculty of Social Science and Humanities at University of Ontario Institute of Technology, states that Lippmann's thesis remains equally relevant in the context of current models of self-management, particularly for privacy [3]. In the previous post, Scott Mason and I had looked at the limitations of a 'notice and consent' regime for privacy governance. Having established the deficiencies of the existing framework for data protection, I will now look at some of the alternatives proposed that may serve to address these issues.

In this article, I will look at paternalistic solutions posed as alternatives to the privacy self-management regime. I will look at theories of paternalism and libertarianism in the context of privacy and with reference to the works of some of the leading philosophers on jurisprudence and political science. The paper will attempt to clarify the main concepts and the arguments put forward by both the proponents and opponents of privacy paternalism. The first alternative solution draws on Anita Allen's thesis in her book, *Unpopular Privacy* [4], which deals with the questions whether individuals have a moral obligation to protect their own privacy. Allen expands the idea of rights to protect one's own self interests and duties towards others to the notion that we may have certain duties not only towards others but also towards ourselves because of their overall impact on the society. In the next section, we will look at the idea of 'libertarian paternalism' as put forth by Cass Sunstein and Richard Thaler [5] and what its impact could be on privacy governance.

Paternalism

Gerald Dworkin, Professor Emeritus at University of California, Davis, defines paternalism as "interference of a state or an individual with another person, against their will, and defended or motivated by a claim that the person interfered with will be better off or protected from harm" [6]. Any act of paternalism will involve some limitation on the autonomy of the subject of the regulation usually without the consent of the subject, and premised on the belief that such act shall either improve the welfare of the subject or prevent it from diminishing [7]. Seana Shiffrin, Professor of Philosophy and Pete Kameron Professor of Law and Social Justice at UCLA, takes a broader view of paternalism and includes within its scope not only matters which are aimed at improving the subject's welfare, but also the replacement of the subject's judgement about matters which may otherwise have lied legitimately within the subject's control [8]. In that sense, Shiffrin's view is interesting for it dispenses with both the requirement for active interference, and such act being premised on the subject's well-being.

The central premise of John Stuart Mill's *On Liberty* is that the only justifiable purpose to exert power over the will of an individual is to prevent harm to others. "His own good, either physical or moral," according to Mill, "is not a sufficient warrant." However, various scholars over the years have found Mill's absolute prohibition problematic and support some degree of paternalism. John Rawls' Principle of Fairness, for instance has been argued to be inherently paternalistic. If one has to put it in a nutshell, the aspect about paternalism

that makes it controversial is that it involves coercion or interference, which in any theory of normative ethics or political science needs to be justified based on certain identified criteria. Staunch opponents of paternalism believe that this justification can never be met. Most scholars however, do not argue that all forms of paternalism are untenable and the bulk of scholarship on paternalism is devoted to formulating the conditions under which this justification is satisfied.

Paternalism interferes with self-autonomy in two ways according to Peter de Marneffe, the Professor of Philosophy at the School of Historical, Philosophical and Religious Studies, Arizona State University [9]. The first is the prohibition principle, under which a person's autonomy is violated by being prohibited from making a choice. The second is the opportunity principle which undermines the autonomy of a person by reducing his opportunities to make a choice. Both the cases should be predicated upon a finding that the paternalistic act will lead to welfare or greater autonomy. According to de Marneffe, there are three conditions under which such acts of paternalism are justified - the benefits of welfare should be substantial, evident and must outweigh the benefits of self-autonomy [10].

There are two main strands of arguments made against paternalism [11]. The first argues that interference with the choices of informed adults will always be an inferior option to letting them decide for themselves, as each person is the 'best judge' of his or her interests. The second strand does not engage with the question about whether paternalism can make better decisions about individuals, but states that any benefit derived from the paternalist act is outweighed by the harm of violation of self-autonomy. Most proponents of soft-paternalism build on this premise by trying to demonstrate that not all paternalistic acts violate self-autonomy. There are various forms of paternalism that we do not question despite them interfering with our autonomy - seat belt laws and restriction of tobacco advertising being a few of them. If we try to locate arguments for self-autonomy in the Kantian framework, it refers not just to the ability to do what one chooses, but to rational self-governance [12]. This theory automatically "opens the door for justifiable paternalism" [13]. In this paper, I assume that certain forms of paternalism are justified. In the remaining two section, I will look at two different theories advocating greater paternalism in the context of privacy governance and try to examine the merits and issues with such measures.

A Moral Obligation to Protect One's Privacy

Modest Paternalism

In her book, *Unpopular Privacy* [14], Anita Allen states that enough emphasis is not placed by people on the value of privacy. The right of individuals to exercise their free will and under the 'notice and consent' regime, give up their rights to privacy as they deem fit is, according to her, problematic. The data protection law in most jurisdictions, is designed to be largely value-neutral in that it does not sit on judgement on what is the nature of information that is being revealed and how the collector uses it. Its primary emphasis is on providing the data subject with information about the above and allowing him to make informed decisions. In my previous post, Scott Mason and I had discussed that with online connectivity becomes increasingly important to participation in modern life, the choice to withdraw completely is becoming less and less of a genuine option [15]. Lamenting that people put little emphasis on privacy and often give away information which, upon retrospection and due consideration, they would feel, they ought not have disclosed, Allen proposes what she calls 'modest paternalism' in which regulations mandate that individuals do not waive their privacy is certain limited circumstances.

Allen acknowledges the tension between her arguments in favor of paternalism and her avowed support for the liberal ideals of autonomy and that government interference should be limited, to the extent possible. However, she tries to make a case for greater paternalism in the context of privacy. She begins by categorizing privacy as a "primary good" essential for "self respect, trusting relationships, positions of responsibility and other forms of

flourishing." In another article, Allen states that this "technophilic generation appears to have made disclosure the default rule of everyday life" [16]. Relying on various anecdotes and examples of individuals' disregard for privacy, she argues that privacy is so "neglected in contemporary life that democratic states, though liberal and feminist, could be justified in undertaking a rescue mission that includes enacting paternalistic privacy laws for the benefit of un-eager beneficiaries." She does state that in most cases it may be more advantageous to educate and incentivise individuals towards making choices that favor greater privacy protection. However, in exceptional cases, paternalism would be justified as a tool to ensure greater privacy.

A Duty Towards Oneself

In an article for the Harvard Symposium on Privacy in 2013, Allen states that laws generally provide a framework built around rights of individuals that enable self-protection and duties towards others. G A Cohen describes Robert Nozick's views which represents this libertarian philosophy as follows: "The thought is that each person is the morally rightful owner of himself. He possesses over himself, as a matter of moral right, all those rights that a slaveholder has over a chattel slave as a matter of legal right, and he is entitled, morally speaking, to dispose over himself in the way such a slaveholder is entitled, legally speaking, to dispose over his slave" [17]. As per the libertarian philosophy espoused by Nozick, everyone is licensed to abuse themselves in the same manner slaveholders abused their slaves.

Allen asks the question whether there is a duty towards oneself and if such a duty exists, should it be reflected in policy or law. She accepts that a range of philosophers consider the idea of duties to oneself as illogical or untenable [18]. Allen, however relies on the works of scholars such as Lara Denis, Paul Eisenberg and Daniel Kading who have located such a duty. She develops a schematic of two kinds of duties - first order duties that requires we protect ourselves for the sake of others, and second order, derivative duties that we protect ourself. Through the essay, she relies on the Kantian framework of categorical imperative to build the moral thrust of her arguments. Kantian view of paternalism would justify those acts which interfere with an individual's autonomy in order to prevent her from exercising her autonomy irrationally, and draw her towards rational end that agree with her conception of good [19]. However, Allen goes one step further and she locates the genesis for duties to both others (perfect duties) and oneself (imperfect duties) in the categorical imperative . Her main thesis is that there are certain situations where we have a moral duty to protect our own privacy where failure to do so would have an impact on either specific others or the society, at large.

Issues

Having built this interesting and somewhat controversial premise, Allen does not sufficiently expand upon it to present a nuanced solution. She provides a number of anecdotes but does not formulate any criteria for when privacy duties could be self-regarding. Her test for what kinds of paternalistic acts are justified is also extremely broad. She argues for paternalism where is protects privacy rights that "enhance liberty, liberal ways of life, well-being and expanded opportunity." She does not clearly define the threshold for when policy should move from incentives to regulatory mandate nor does she elaborate upon what forms paternalism would both serve the purpose of protecting privacy as well as ensuring that there is no unnecessary interference with the rights of individual [20].

Nudge and Libertarian Paternalism

What is Nudge?

In 2006, Richard Thaler and Cass Sunstein published their book *Nudge: Improving decisions about health, wealth and happiness* [21]. The central thesis of the book is that in order to make most of decisions, we rely on a menu of options made available to us and the order

and structure of choices is characterised by Thaler and Sunstein as "choice architecture." According to them, the choice architecture has a significant impact on the choices that we make. The book looks at examples from a food cafeteria, the position of restrooms and how whether the choice is to opt-in or opt-out influences the retirement plans that were chosen. This choice architecture influences our behavior without coercion or a set of incentives, as conventional public policy theory would have us expect. The book draws on work done by cognitive scientists such as Daniel Kahneman [22] and Amos Tversky [23] as well as Thaler's own research in behavioral economics [24]. The key takeaway from cognitive science and behavioral economics used in this book is that choice architecture influences our actions in anticipated ways and leads to predictably irrational behavior. Thaler and Sunstein believe that this presents a great potential for policy makers. They can tweak the choice architecture in their specific domains to influence the decisions made by its subjects and nudge them towards behavior that is beneficial to them and/or the society.

The great attraction of the argument made by Thaler and Sunstein is that it offers a compromise between forbearance and mandatory regulation. If we identify the two ends of the policy spectrum as - a) paternalists who believe in maximum interference through legal regulations that coerce behavior to meet the stated goals of the policy, and b) libertarians who believe in the free market theory that relies on the individuals making decisions in their best interests, 'nudging' falls somewhere in the middle, leading to the oxymoronic yet strangely apt phrase, "libertarian paternalism." The idea is to design choices in such as way that they influence decision-making so as to increase individual and societal welfare. In his book, *The Laws of Fear*, Cass Sunstein argues that the anti-paternalistic position is incoherent as "there is no way to avoid effects on behavior and choices."

The proponents of libertarian paternalism refute the commonly posed question about who decides the optimal and desirable results of choice architecture, by stating that this form of paternalism does not promote a perfectionist standard of welfare but an individualistic and subjective standard. According to them, choices are not prohibited, cordoned off or made to carry significant barriers. However, it is often difficult to conclude what it is that is better for the welfare of people, even from their own point of view. The claim that nudges lead to choices that make them better off by their own standards seems more and more untenable. What nudges do is lead people towards certain broad welfare which the choice-architects believe make the lives of people better in the longer term [25].

How Nudges Could Apply to Privacy?

Our previous post echoes the assertion made by Thaler and Sunstein that the traditional rational choice theory that assumes that individuals will make rationally optimal choices in their self interest when provided with a set of incentives and disincentives, is largely a fiction. We have argued that this assertion holds true in the context of privacy protection principles of notice and informed consent. Daniel Solove has argued that insights from cognitive science, particularly using the theory of nudge would be an acceptable compromise between the inefficacy of privacy self-management and the dangers of paternalism [26]. His rationale is that while nudges influence choice, they are not overly paternalistic in that they still give the individual the option of making choices contrary to those sought by the choice architecture. This is an important distinction and it demonstrates that 'nudging' is less coercive than how we generally understand paternalistic policies.

One of the nudging techniques which makes a lot of sense in the context of the data protection policies is the use of defaults. It relies on the oft-mentioned status quo bias [27]. This is mentioned by Thaler and Sunstein with respect to encouraging retirement savings plans and organ donation, but would apply equally to privacy. A number of data collectors have maximum disclosure as their default settings and effort in understanding and changing these settings is rarely employed by users. A rule which mandates that data collectors set optimal defaults that ensure that the most sensitive information is subjected to least degree of disclosure unless otherwise chosen by the user, will ensure greater privacy protection.

Ryan Calo and Dr. Victoria Groom explored an alternative to the traditional notice and consent regime at the Centre of Internet and Society, Stanford University [28]. They conducted a two-phase experimental study. In the first phase, a standard privacy notice was compared with a control condition and a simplified notice to see if improving the readability impacted the response of users. In the second phase, the notice was compared with five notices strategies, out of which four were intended to enhance privacy protective behavior and one was intended to lower it. Shara Monteleone and her team used a similar approach but with a much larger sample size [29]. One of the primary behavioral insights used was that when we do repetitive activities including accepting online terms and conditions or privacy notices, we tend to use our automatic or fast thinking instead to reflective or slow thinking [30]. Changing them requires leveraging the automatic behavior of the individuals.

Alessandro Acquisti, Professor of Information Technology and Public Policy at the Heinz College, Carnegie Mellon University, has studied the application of methodologies from behavioral economics to investigate privacy decision-making [31]. He highlights a variety of factors that distort decision-making such as - "inconsistent preferences and frames of judgment; opposing or contradictory needs (such as the need for publicity combined with the need for privacy); incomplete information about risks, consequences, or solutions inherent to provisioning (or protecting) personal information; bounded cognitive abilities that limit our ability to consider or reflect on the consequences of privacy-relevant actions; and various systematic (and therefore predictable) deviations from the abstractly rational decision process." Acquisti looks at three kinds of policy solutions taking the example of social networking sites collecting sensitive information- a) hard paternalistic approach which ban making visible certain kind of information on the site, b) a usability approach that entails designing the system in way that is most intuitive and easy for users to decide whether to provide the information, c) a soft paternalistic approach which seeks to aid the decision-making by providing other information such as how many people would have access to the information, if provided, and set defaults such that the information is not visible to others unless explicitly set by the user. The last two approaches are typically cited as examples of nudging approaches to privacy.

Another method is to use tools that lead to decreased disclosure of information. For example, tools like Social Media Sobriety Test [32] or Mail Goggles [33] serve to block the sites during certain hours set by user during which one expects to be at their most vulnerable, and the online services are blocked unless the user can pass a dexterity examination [34]. Rebecca Belabako and her team are building privacy enhanced tools for Facebook and Twitter that will provide greater nudges in restricting who they share their location on Facebook and restricting their tweets to smaller group of people [35]. Ritu Gulia and Dr. Sapna Gambhir have suggested nudges for social networking websites that randomly select pictures of people who will have access to the information to emphasise the public or private setting of a post [36]. These approaches try to address the myopia bias where we choose immediate access to service over long term privacy harms.

The use of nudges as envisioned in the examples above is in some ways an extension of already existing research which advocates a design standard that makes the privacy notices more easily intelligible [37]. However, studies show only an insignificant improvement by using these methods. Nudging, in that sense goes one step ahead. Instead of trying to make notices more readable and enable informed consent, the design standard will be intended to simply lead to choices that the architects deem optimal.

Issues with Nudging

One of the primary justifications that Thaler and Sunstein put forward for nudging is that the choice architecture is ubiquitous. The manner in which option are presented to us impact how we make decision whether it was intended to do so or not, and that there is no such thing a neutral architecture. This inevitability, according to them, makes a strong case for nudging people towards choices that will lead to their well-being. However, this assessment

does not support the arguments made by them that libertarian paternalism nudges people towards choices from their own point of view. It is my contention that various examples of libertarian paternalism, as put forth by Thaler and Sunstein, do in fact interfere with our self-autonomy as the choice architecture leads us not to options that we choose for ourselves in a fictional neutral environments, but to those options that the architects believe are good for us. This substitution of judgment would satisfy the definition by Seana Shiffron. Second, the fact that there is no such things as a neutral architecture, is by itself, not justification enough for nudging. If we view the issue only from the point of view of normative ethics, assuming that coercion and interference are undesirable, intentional interference is much worse than unintentional interference.

However, there are certain nudges that rely primarily on providing information, dispensing advice and rational persuasion [38]. The freedom of choice is preserved in these circumstances. Libertarians may argue that even these circumstances the shaping of choice is problematic. This issue, J S Blumenthal-Barby argues, is adequately addressed by the publicity condition, a concept borrowed by Thaler and Sunstein from John Rawls [39]. The principle states that officials should never use a technique they would be uncomfortable defending to the public; nudging is no exception. However, this seems like a simplistic solution to a complex problem. Nudges are meant to rely on inherent psychological tendencies, leveraging the theories about automatic and subconscious thinking as described by Daniel Kahneman in his book, "Thinking Fast, Thinking Slow [40]." In that sense, while transparency is desirable it may not be very effective.

Other commentators also note that while behavioral economics can show why people make certain decisions, it may not be able to reliably predict how people will behave in different circumstances. The burden of extrapolating the observations into meaningful nudges may prove to be too heavy [41]. However, the most oft-quoted criticism of nudging is that it will rely on officials to formulate the desired goals towards which the choice architecture will lead us [42]. The judgments of these officials could be flawed and subject to influence by large corporations [43]. These concerns echo the best judge argument made against all forms of paternalism, mentioned earlier in this essay. J S Blumenthal-Barby, Assistant Professor at the Center for Medical Ethics and Health Policy, Baylor College of Medicine, also examines the claim that the choice architects will be susceptible to the same biases while designing the choice environment [44]. His first argument in response to this is that experts who extensively study decision-making may be less prone to these errors. Second, he argues that even with errors and biases, a choice architecture which attempts to the rights the wrongs of a random and unstructured choice environment is a preferable option [45].

Conclusion

Most libertarians will find the notion that individuals are prevented from sharing some information about themselves problematic. Anita Allen's idea about self-regarding duties is at odds how we understand rights and duties in most jurisdictions. Her attempt to locate an ethical duty to protect one's privacy, while interesting, is not backed by a formulation of how such a duty would work. While she relies largely on an Kantian framework, her definition of paternalism, as can be drawn from her writing is broader than that articulated by Kant himself. On the other hand, Thaler and Sunstein's book Nudge and related writings by them do attempt to build a framework of how nudging would work and answer some questions they anticipate would be raised against the idea of libertarian paternalism.

By and large, I feel that, Thaler and Sunstein's idea of libertarian paternalism could be justified in the context of privacy and data protection governance. It would be fair to say the first two conditions of de Marneffe under which such acts of paternalism are justified [46] are largely satisfied by nudges that ensures greater privacy protection. If nudges can ensure greater privacy protection, its benefits are both substantial and evident. However, the larger question is whether these purported benefits outweigh the costs of loss of self-autonomy. Given the numerous ways in which the 'notice and consent' framework is

ineffective and leads to very little informed consent, it can be argued that there is little exercise of autonomy, to begin with, and hence, the loss of self-autonomy is not substantial. Some of the conceptual issues which doubt the ability of nudges to solve complex problems remain unanswered and we will have to wait for more analysis by both cognitive scientists and policy-makers. However, given the growing inefficacy of the existing privacy protection framework, it would be a good idea of begin using some insights from cognitive science and behavioral economics to ensure greater privacy protection.

The current value-neutrality of data protection law with respect of the kind of data collected and its use, and its complete reliance on the data subject to make an informed choice is, in my opinion, an idea that has run its course. Rather than focussing solely on the controls at the stage of data collection, I believe we need a more robust theory of how to govern the subsequent uses of data. This will is the focus of the next part of this series in which I will look at the greater use of risk-based approach to privacy protection.

ENDNOTES

[1] With invaluable inputs from Scott Mason.

[2] Walter Lippmann, The Phantom Public, Transaction Publishers, 1925.

[3] Jonathan Obar, Big Data and the Phantom Public: Walter Lippmann and the fallacy of data privacy self management, Big Data and Society, 2015, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2239188

[4] Anita Allen, Unpopular Privacy: What we must hide?, Oxford University Press USA, 2011.

[5] Richard Thaler and Cass Sunstein, Nudge, Improving decisions about health, wealth and happinessYale University Press, 2008.

[6] http://plato.stanford.edu/entries/paternalism/

[7] Christian Coons and Michael Weber, ed., Paternalism: Theory and Practice; Cambridge University Press, 2013. at 29.

[8] Seana Shiffrin, Paternalism, Unconscionability Doctrine, and Accommodation, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2682745

[9] Peter de Marneffe, Self Sovereignty and Paternalism, from Christian Coons and Michael Weber, ed., Paternalism: Theory and Practice; Cambridge University Press, 2013. at 58.

[10] Id .

[11] Christian Coons and Michael Weber, ed., Paternalism: Theory and Practice; Cambridge University Press, 2013. at 74.

[12] Christian Coons and Michael Weber, ed., Paternalism: Theory and Practice; Cambridge University Press, 2013. at 115.

[13] Ibid at 116.

[14] Anita Allen, Unpopular Privacy: What we must hide?, Oxford University Press USA, 2011.

[15] Janet Vertasi, My Experiment Opting Out of Big Data Made Me Look Like a Criminal, 2014, available at http://time.com/83200/privacy-internet-big-data-opt-out/

[16] Anita Allen, Privacy Law: Positive Theory and Normative Practice, available at http://harvardlawreview.org/2013/06/privacy-law-positive-theory-and-normative-practice/ .

[17] G A Cohen, Self ownership, world ownership and equality, available at http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=3093280

[18] Marcus G. Singer, On Duties to Oneself, available at http://www.jstor.org/stable/2379349?seq=1#page_scan_tab_contents; Kurt Baier, The moral point of view: A rational basis of ethics, available athttps://www.uta.edu/philosophy/faculty/burgess-jackson/Baier,%20The%20Moral%20Point%20of%20View%20%281958%29%20%28Excerpt%20on%20Ethical%20Egoism%29.pdf.

[19] Michael Cholbi, Kantian Paternalism and suicide intervention, from Christian Coons and Michael Weber, ed., Paternalism: Theory and Practice; Cambridge University Press, 2013.

[20] Eric Posner, Liberalism and Concealment, available at https://newrepublic.com/article/94037/unpopular-privacy-anita-allen

[21] Richard Thaler and Cass Sunstein, Nudge, Improving decisions about health, wealth and happinessYale University Press, 2008.

[22] Daniel Kahneman, Thinking, fast and slow, Farrar, Straus and Giroux, 2011.

[23] Daniel Kahneman, Paul Slovic and Amos Tversky, Judgment under uncertainty: heuristics and biases, Cambridge University Press, 1982; Daniel Kahneman and Amos Tversky, Choices, Values and Frames, Cambridge University Press, 2000.

[24] Richard Thaler, Advances in behavioral finance, Russell Sage Foundation, 1993.

[25] Thaler, Sunstein and Balz, Choice Architecture, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1583509.

[26] Daniel Solove, Privacy self-management and consent dilemma, 2013 available at http://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2093&context=faculty_publications

[27] Frederik Borgesius, Behavioral sciences and the regulation of privacy on the Internet, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2513771.

[28] Ryan Calo and Dr. Victoria Groom, Reversing the Privacy Paradox: An experimental study, available at http://ssrn.com/abstract=1993125

[29] Shara Monteleon et al, Nudges to Privacy Behavior: Exploring an alternative approahc to privacy notices, available at http://publications.jrc.ec.europa.eu/repository/bitstream/JRC96695/jrc96695.pdf

[30] Daniel Kahneman, Thinking, fast and slow, Farrar, Straus and Giroux, 2011.

[31] Alessandro Acquisti, Nudging Privacy, available at http://www.heinz.cmu.edu/~acquisti/papers/acquisti-privacy-nudging.pdf

[32] http://www.webroot.com/En_US/sites/sobrietytest/test.php?url=0

[33] http://google.about.com/od/m/g/mail_goggles.htm

[34] Rebecca Balebako et al, Nudging Users towards privacy on mobile devices, available at https://www.andrew.cmu.edu/user/pgl/paper6.pdf.

[35] Id .

[36] Ritu Gulia and Dr. Sapna Gambhir, Privacy and Privacy Nudges for OSNs: A Review, available at http://www.ijircce.com/upload/2014/march/14L_Privacy.pdf

[37] Annie I. Anton et al., Financial Privacy Policies and the Need for Standardization, 2004 available at https://ssl.lu.usi.ch/entityws/Allegati/pdf_pub1430.pdf; Florian Schaub, R. Balebako et al, “A Design Space for effective privacy notices” available at https://www.usenix.org/system/files/conference/soups2015/soups15-paper-schaub.pdf

[38] Daniel Hausman and Bryan Welch argue that these cases are mistakenly characterized as nudges. They believe that nudges do not try to inform the automatic system, but manipulate the inherent cognitive biases. Daniel Hausman and Bryan Welch, Debate: To Nudge or Not to Nudge, Journal of Political Philosophy 18(1).

[39] Ryan Calo, Code, Nudge or Notice, available at

[40] Daniel Kahneman, Thinking, fast and slow, Farrar, Straus and Giroux, 2011.

[41] Evan Selinger and Kyle Powys Whyte, Nudging cannot solve complex policy problems.

[42] Mario J. Rizzo & Douglas Glen Whitman, The Knowledge Problem of New Paternalism, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1310732; Pierre Schlag, Nudge, Choice Architecture, and Libertarian Paternalism, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1585362.

[43] Edward L. Glaeser, Paternalism and Psychology, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=917383.

[44] J S BLumenthal-Barby, Choice Architecture: A mechanism for improving decisions while preserving liberty?, from Christian Coons and Michael Weber, ed., Paternalism: Theory and Practice; Cambridge University Press, 2013.

[45] Id.

[46] According to de Marneffe, there are three conditions under which such acts of paternalism are justified - the benefits of welfare should be substantial, evident and must outweigh the benefits of self-autonomy. Peter de Marneffe, Self Sovereignty and Paternalism, from Christian Coons and Michael Weber, ed., Paternalism: Theory and Practice; Cambridge University Press, 2013. at 58.

New Approaches to Information Privacy – Revisiting the Purpose Limitation Principle

AMBER SINHA

Introduction

Last year, Mukul Rohatgi, the Attorney General of India, called into question existing jurisprudence of the last 50 years on the constitutional validity of the right to privacy [1]. Mohatgi was rebutting the arguments on privacy made against Aadhaar, the unique identity project initiated and implemented in the country without any legislative mandate [2]. The question of the right to privacy becomes all the more relevant in the context of events over the last few years—among them, the significant rise in data collection by the state through various e-governance schemes [3], systematic access to personal data by various wings of the state through a host of surveillance and law enforcement initiatives launched in the last decade [4] the multifold increase in the number of Indians online, and the ubiquitous collection of personal data by private parties [5].

These developments have led to a call for a comprehensive privacy legislation in India and the adoption of the National Privacy Principles as laid down by the Expert Committee led by Justice AP Shah [6]. There are privacy-protection legislation currently in place such as the Information Technology Act, 2000 (IT Act), which was enacted to govern digital content and communication and provide legal recognition to electronic transactions. This legislation has provisions that can safeguard—and dilute—online privacy. At the heart of the data protection provisions in the IT Act lies section 43A and the rules framed under it, i.e., Reasonable security practices and procedures and sensitive personal data information [7]. Section 43A mandates that body corporates who receive, possess, store, deal, or handle any personal data to implement and maintain 'reasonable security practices', failing which, they are held liable to compensate those affected. Rules drafted under this provision also mandated a number of data protection obligations on corporations such the need to seek consent before collection, specifying the purposes of data collection, and restricting the use of data to such purposes only. There have been questions raised about the validity of the Section 43A Rules as they seek to do much more than mandate in the parent provisions, Section 43A— requiring entities to maintain reasonable security practices.

Privacy as Control?

Even setting aside the issue of legal validity, the kind of data protection framework envisioned by Section 43A rules is proving to be outdated in the context of how data is now being collected and processed. The focus of Section 43 A Rules—as well as that of draft privacy legislations in India [8] —is based on the idea of individual control. Most apt is Alan Westin's definition of privacy: "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to other" [9]. Westin and his followers rely on the normative idea of "informational self-determination", the notion of a pure, disembodied, and atomistic self, capable of making rational and isolated choices in order to assert complete control over personal information. More and more this has proved to be a fiction especially in a networked society.

Much before the need for governance of information technologies had reached a critical mass in India, Western countries were already dealing with the implications of the use of these technologies on personal data. In 1973, the US Department of Health, Education and Welfare appointed a committee to address this issue, leading to a report called 'Records, Computers and Rights of Citizens' [10]. The Committee's mandate was to "explore the impact of computers on record keeping about individuals and, in addition, to inquire into, and make

recommendations regarding, the use of the Social Security number." The Report articulated five principles which were to be the basis of fair information practices: transparency; use limitation; access and correction; data quality; and security. Building upon these principles, the Committee of Ministers of the Organization for Economic Cooperation and Development (OECD) arrived at the Guidelines on the Protection of Privacy and Transborder Flows of Personal Data in 1980 [11]. These principles— Collection Limitation, Data Quality, Purpose Specification, Use Limitation, Security Safeguards, Openness, Individual Participation and Accountability—are what inform most data protection regulations today including the APEC Framework, the EU Data Protection Directive, and the Section 43A Rules and Justice AP Shah Principles in India.

Fred Cate describes the import of these privacy regimes as such:

"All of these data protection instruments reflect the same approach: tell individuals what data you wish to collect or use, give them a choice, grant them access, secure those data with appropriate technologies and procedures, and be subject to third-party enforcement if you fail to comply with these requirements or individuals' expressed preferences" [12].

This is in line with Alan Westin's idea of privacy exercised through individual control. Therefore the focus of these principles is on empowering the individuals to exercise choice, but not on protecting individuals from harmful or unnecessary practices of data collection and processing. The author of this article has earlier written [13] about the sheer inefficacy of this framework which places the responsibility on individuals. Other scholars like Daniel Solove [14], Jonathan Obar [15] and Fred Cate [16] have also written about the failure of traditional data protection practices of notice and consent. While these essays dealt with the privacy principles of choice and informed consent, this paper will focus on the principles of purpose limitation.

Purpose Limitation and Impact of Big Data

The principles of purpose limitation or purpose specification seeks to ensure the following four objectives:

a. Personal information collected and processed should be adequate and relevant to the purposes for which they are processed.

b. The entities collect, process, disclose, make available, or otherwise use personal information only for the stated purposes.

c. In case of change in purpose, the data's subject needs to be informed and their consent has to be obtained.

d. After personal information has been used in accordance with the identified purpose, it has to be destroyed as per the identified procedures.

The purpose limitation along with the data minimisation principle—which requires that no more data may be processed than is necessary for the stated purpose—aim to limit the use of data to what is agreed to by the data subject. These principles are in direct conflict with new technology which relies on ubiquitous collection and indiscriminate uses of data. The main import of Big Data technologies on the inherent value in data which can be harvested not by the primary purposes of data collection but through various secondary purposes which involve processing of the data repeatedly [17]. Further, instead to destroying the data when its purpose has been achieved, the intent is to retain as much data as possible for secondary uses. Importantly, as these secondary uses are of an inherently unanticipated nature, it becomes impossible to account for it at the stage of collection and providing the choice to the data subject.

Followers of the discourse on Big Data would be well aware of its potential impacts on privacy. De-identification techniques to protect the identities of individuals in dataset face a threat from an increase in the amount of data available either publicly or otherwise to

a party seeking to reverse-engineer an anonymised dataset to re-identify individuals [18]. Further, Big Data analytics promise to find patterns and connections that can contribute to the knowledge available to the public to make decisions. What is also likely is that it will lead to revealing insights about people that they would have preferred to keep private [19]. In turn, as people become more aware of being constantly profiled by their actions, they will self-regulate and 'discipline' their behaviour. This can lead to a chilling effect [20]. Meanwhile, Big Data is also fuelling an industry that incentivises businesses to collect more data, as it has a high and growing monetary value. However, Big Data also promises a completely new kind of knowledge that can prove to be revolutionary in fields as diverse as medicine, disaster-management, governance, agriculture, transport, service delivery, and decision-making [21]. As long as there is a sufficiently large and diverse amount of data, there could be invaluable insights locked in it, accessing which can provide solutions to a number of problems. In light of this, it is important to consider what kind of regulatory framework is most suitable which could facilitate some of the promised benefits of Big Data and at the same time mitigate its potential harm. This, coupled with the fact that the existing data protection principles have, by most accounts, run their course, makes the examination of alternative frameworks even more important. This article will examine some alternate proposals made to the existing framework of purpose limitation below.

Harms-Based Approach

Some scholars like Fred Cate [22] and Daniel Solove [23] have argued that there is a need for the primary focus of data protection law to move from control at the stage of data collection to actual use cases. In his article on the failure of Fair Information Practice Principles [24], Cate puts forth a proposal for 'Consumer Privacy Protection Principles.' Cate envisions a more interventionist role of the data protection authorities by regulating information flows when required, in order to protect individuals from risky or harmful uses of information. Cate's attempt is to extend the principles of consumer protection law of prevention and remedy of harms.

In a re-examination of the OECD Privacy Principles, Cate and Viktor Mayer Schöemberger attempt to discard the use of personal data to only purposes specified. They felt that restricting the use of personal to only specified purposes could significantly threaten various research and beneficial uses of Big Data. Instead of articulating a positive obligations of what personal data collected could be used for, they attempt to arrive at a negative obligation of use-cases prevented by law. Their working definition of the Use specification principle broaden the scope of use cases by only preventing use of data "if the use is fraudulent, unlawful, deceptive or discriminatory; society has deemed the use inappropriate through a standard of unfairness; the use is likely to cause unjustified harm to the individual; or the use is over the well-founded objection of the individual, unless necessary to serve an over-riding public interest, or unless required by law" [25].

While most standards in the above definition have established understanding in jurisprudence, the concept of unjustifiable harm is what we are interested in. Any theory of harms-based approach goes back to John Stuart Mill's dictum that the only justifiable purpose to exert power over the will of an individual is to prevent harm to others. Therefore, any regulation that seeks to control or prevent autonomy of individuals (in this case, the ability of individuals to allow data collectors to use their personal data, and the ability of data collectors to do so, without any limitation) must clearly demonstrate the harm to the individuals in question.

Fred Cate articulates the following steps to identify tangible harm and respond to its presence [26]:

a. **Focus on Use** — Actual use of the data should be considered, not mere possession. The assumption is that the collection, possession, or transfer of information do not significantly harm people, rather it is the use of information following such collection, possession, or transfer.

b. **Proportionality** — Any regulatory measure must be proportional to the likelihood and severity of the harm identified.

c. **Per se Harmful Uses** — Uses which are always harmful must be prohibited by law

d. **Per se not Harmful Uses** — If uses can be considered inherently not harmful, they should not be regulated.

e. **Sensitive Uses** — In case where the uses are not per se harmful or not harmful, individual consent must be sought for using that data for those purposes.

The proposal by Cate argues for what is called a 'use based system', which is extremely popular with American scholars. Under this system, data collection itself is not subject to restrictions; rather, only the use of data is regulated. This argument has great appeal for both businesses who can reduce their overheads significantly if consent obligations are done away with as long as they use the data in ways which are not harmful, as well as critics of the current data protection framework which relies on informed consent. Lokke Moerel explains the philosophy of 'harms based approach' or 'use based system' in United States by juxtaposing it against the 'rights based approach' in Europe [27]. In Europe, rights of individuals with regard to processing of their personal data is a fundamental human right and therefore, a precautionary principle is followed with much greater top-down control upon data collection. However, in the United States, there is a far greater reliance on market mechanisms and self-regulating organisations to check inappropriate processing activities, and government intervention is limited to cases where a clear harm is demonstrable [28].

Continuing research by the Centre for Information Policy Leadership under its Privacy Risk Framework Project looks at a system of articulating what harms and risks arising from use of collected data. They have arrived a matrix of threats and harms. Threats are categorised as — a) inappropriate use of personal information and b) personal information in the wrong hands. More importantly for our purposes, harms are divided into: a) tangible harms which are physical or economic in nature (bodily harm, loss of liberty, damage to earning power and economic interests); b) intangible harms which can be demonstrated (chilling effects, reputational harm, detriment from surveillance, discrimination and intrusion into private life); and c) societal harm (damage to democratic institutions and loss of social trust) [29]. For any harms-based system, a matrix like above needs to emerge clearly so that regulation can focus on mitigating practices leading to the harms.

Legitimate Interests

Lokke Moerel and Corien Prins, in their article "Privacy for Homo Digitalis – Proposal for a new regulatory framework for data protection in the light of Big Data and Internet of Things" [30] use the ideal of responsive regulation which considers empirically observable practices and institutions while determining the regulation and enforcement required. They state that current data protection frameworks—which rely on mandating some principles of how data has to be processed—is exercised through merely procedural notification and consent requirements. Further, Moerel and Prins feel that data protection law cannot only involve a consideration of individual interest but also needs to take into account collective interest. Therefore, the test must be a broader assessment than merely the purpose limitation articulating the interests of the parties directly involved, but whether a legitimate interest is achieved.

Legitimate interest has been put forth as an alternative to the purpose limitation. Legitimate is not a new concept and has been a part of the EU Data Protection Directive and also finds a place in the new General Data Protection Regulation. Article 7 (f) of the EU Directive [31] provided for legitimate interest balanced against the interests or fundamental rights and freedoms of the data subject as the last justifiable reason for use of data. Due to confusion in its interpretation, the Article 29 Working Party, in 2014 [32], looked into the role of legitimate interest and arrived at the following factors to determine the presence of a legitimate interest— a) the status of the individual (employee, consumer, patient) and the controller

(employer, company in a dominant position, healthcare service); b) the circumstances surrounding the data processing (contract relationship of data subject and processor); c) the legitimate expectations of the individual.

Federico Ferretti has criticised the legitimate interest principle as vague and ambiguous. The balancing of legitimate interest in using the data against fundamental rights and freedoms of the data subject gives the data controllers some degree of flexibility in determining whether data may be processed; however, this also reduces the legal certainty that data subject have of their data not being used for purposes they have not agreed to [33]. However, it is this paper's contention that it is not the intent of the legitimate interest criteria but the lack of consensus on its application which creates an ambiguity. Moerel and Prins articulate a test for using legitimate interest which is cognizant of the need to use data for the purpose of Big Data processing, as well as ensuring that the rights of data subjects are not harmed.

As demonstrated earlier, the processing of data and its underlying purposes have become exceedingly complex and the conventional tool to describe these processes 'privacy notices' are too lengthy, too complex and too profuse in numbers to have any meaningful impact [34]. The idea of information self-determination, as contemplated by Westin in American jurisprudence, is not achieved under the current framework. Moerel and Prins recommend five factors [35] as relevant in determining the legitimate interest. Of the five, the following three are relevant to the present discussion:

a. **Collective Interest** — A cost-benefit analysis should be conducted, which examines the implications for privacy for the data subjects as well as the society, as a whole.

b. **The nature of the data** — Rather than having specific categories of data, the nature of data needs to be assessed contextually to determine legitimate interest.

c. **Contractual relationship and consent not independent grounds** — This test has two parts. First, in case of contractual relationship between data subject and data controller: the more specific the contractual relationship, the more restrictions apply to the use of the data. Second, consent does not function as a separate principle which, once satisfied, need not be revisited. The nature of the consent (opportunities made available to data subject, opt in/opt out, and others) will continue to play a role in determining legitimate interest.

Conclusion

Replacing the purpose limitation principles with a use-based system as articulated above poses the danger of allowing governments and the private sector to carry out indiscriminate data collection under the blanket guise that any and all data may be of some use in the future. The harms-based approach has many merits and there is a stark need for more use of risk assessments techniques and privacy impact assessments in data governance. However, it is important that it merely adds to the existing controls imposed at data collection, and not replace them in their entirety. On the other hand, the legitimate interests principle, especially as put forth by Moerel and Prins, is more cognizant of the different factors at play — the inefficacy of existing purpose limitation principles, the need for businesses to use data for purposes unidentified at the stage of collection, and the need to ensure that it is not misused for indiscriminate collection and purposes. However, it also poses a much heavier burden on data controllers to take into account various factors before determining legitimate interest. If legitimate interest has to emerge as a realistic alternative to purpose limitation, there needs to be greater clarity on how data controllers must apply this principle.

ENDNOTES

[1] Prachi Shrivastava, "Privacy not a fundamental right, argues Mukul Rohatgi for Govt as Govt affidavit says otherwise," Legally India, Jyly 23, 2015, http://www.legallyindia.com/Constitutional-law/privacy-not-a-fundamental-right-argues-mukul-rohatgi-for-govt-as-govt-affidavit-says-otherwise.

[2] Rebecca Bowe, "Growing Mistrust of India's Biometric ID Scheme," Electronic Frontier Foundation, May 4, 2012, https://www.eff.org/deeplinks/2012/05/growing-mistrust-india-biometric-id-scheme.

[3] Lisa Hayes, "Digital India's Impact on Privacy: Aadhaar numbers, biometrics, and more," Centre for Democracy and Technology, January 20, 2015, https://cdt.org/blog/digital-indias-impact-on-privacy-aadhaar-numbers-biometrics-and-more/.

[4] "India's Surveillance State," Software Freedom Law Centre, http://sflc.in/indias-surveillance-state-our-report-on-communications-surveillance-in-india/.

[5] "Internet Privacy in India," Centre for Internet and Society, http://cis-india.org/telecom/knowledge-repository-on-internet-access/internet-privacy-in-india.

[6] Vivek Pai, "Indian Government says it is still drafting privacy law, but doesn't give timelines," Medianama, May 4, 2016, http://www.medianama.com/2016/05/223-government-privacy-draft-policy/.

[7] Information Technology (Intermediaries Guidelines) Rules, 2011, http://deity.gov.in/sites/upload_files/dit/files/GSR314E_10511%281%29.pdf.

[8] Discussion Points for the Meeting to be taken by Home Secretary at 2:30 pm on 7-10-11 to discuss the drat Privacy Bill, http://cis-india.org/internet-governance/draft-bill-on-right-to-privacy.

[9] Alan Westin, Privacy and Freedom (New York: Atheneum, 2015).

[10] US Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computers and the Rights of Citizens, http://www.justice.gov/opcl/docs/rec-com-rights.pdf.

[11] OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

[12] Fred Cate, "The Failure of Information Practice Principles," in Consumer Protection in the Age of the Information Economy, ed. Jane K. Winn (Burlington: Aldershot, Hants, England, 2006) http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1156972.

[13] Amber Sinha and Scott Mason, "A Critique of Consent in Informational Privacy," Centre for Internet and Society, January 11, 2016, http://cis-india.org/internet-governance/blog/a-critique-of-consent-in-information-privacy.

[14] Daniel Solove, "Privacy self-management and consent dilemma," Harvard Law Review 126, (2013): 1880.

[15] Jonathan Obar, "Big Data and the Phantom Public: Walter Lippmann and the fallacy of data privacy self management," Big Data and Society 2(2), (2015), doi: 10.1177/2053951715608876.

[16] Supra Note 12.

[17] Supra Note 14.

[18] Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization" available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006; Arvind Narayanan and Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Datasets" available at https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf.

[19] D. Hirsch, "That's Unfair! Or is it? Big Data, Discrimination and the FTC's Unfairness Authority," Kentucky Law Journal, Vol. 103, available at: http://www.kentuckylawjournal.org/wp-content/uploads/2015/02/103KyLJ345.pdf

[20] A Marthews and C Tucker, "Government Surveillance and Internet Search Behavior", available at http://ssrn.com/abstract=2412564; Danah Boyd and Kate Crawford, "Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon", Information, Communication & Society, Vol. 15, Issue 5, (2012).

[21] Scott Mason, "Benefits and Harms of Big Data", Centre for Internet and Society, available at http://cis-india.org/internet-governance/blog/benefits-and-harms-of-big-data#_ftn37.

[22] Cate, "The Failure of Information Practice Principles."

[23] Solove, "Privacy self-management and consent dilemma," 1882.

[24] Cate, "The Failure of Information Practice Principles."

[25] Fred Cate and Viktor Schoenberger, "Notice and Consent in a world of Big Data," International Data Privacy Law 3(2), (2013): 69.

[26] Solove, "Privacy self-management and consent dilemma," 1883.

[27] Lokke Moerel, "Netherlands: Big Data Protection: How To Make The Draft EU Regulation On Data Protection Future Proof", Mondaq, March 11. 2014, http://www.mondaq.com/x/298416/data+protection/Big+Data+Protection+How+To+Make+The+Dra%20ft+EU+Regulation+On+Data+Protection+Future+Proof%20al%20Lecture.

[28] Moerel, “Netherlands: Big Data Protection.”

[29] Centre for Information Policy Leadership, “A Risk-based Approach to Privacy: Improving Effectiveness in Practice,” Hunton and Williams LLP, June 19, 2014, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/white_paper_1-a_risk_based_approach_to_privacy_improving_effectiveness_in_practice.pdf.

[30] Lokke Moerel and Corien Prins, “Privacy for Homo Digitalis: Proposal for a new regulatory framework for data protection in the light of Big Data and Internet of Things”, Social Science Research Network, May 25, 2016, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2784123.

[31] EU Directive 95/46/EC – The Data Protection Directive, https://www.dataprotection.ie/docs/EU-Directive-95-46-EC-Chapter-2/93.htm.

[32] Article 29 Data Protection Working Party, “Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC,” http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf.

[33] Frederico Ferretti, “Data protection and the legitimate interest of data controllers: Much ado about nothing or the winter of rights?,” Common Market Law Review 51(2014): 1-26. http://bura.brunel.ac.uk/bitstream/2438/9724/1/Fulltext.pdf.

[34] Sinha and Mason, “A Critique of Consent in Informational Privacy.”

[35] Moerel and Prins, “Privacy for Homo Digitalis.”

Links to Articles

Legal Landscape for Privacy
cis-india.org/internet-governance/blog/state-of-cyber-security-and-surveillance-in-india.pdf

Nature of Knowledge
cis-india.org/internet-governance/blog/nature-of-knowledge

Benefits and Harms of "Big Data"
cis-india.org/internet-governance/blog/benefits-and-harms-of-big-data

A Review of the Policy Debate around Big Data and Internet of Things
cis-india.org/internet-governance/blog/review-of-policy-debate-around-big-data-and-internet-of-things

Big Data and the Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules 2011
cis-india.org/internet-governance/blog/big-data-and-information-technology-rules-2011

Too Clever by Half: Strengthening India's Smart Cities Plan with Human Rights Protection
cis-india.org/internet-governance/blog/the-wire-march-21-2016-vanya-rakesh-too-clever-by-half-strengthening-indias-smart-cities-plan-with-human-rights-protection

Predictive Policing: What is it, How it Works, and its Legal Implications
cis-india.org/internet-governance/blog/predictive-policing-what-is-it-how-it-works-and-it-legal-implications

Digital Emergency Power: Big Data and Ebola Response in Liberia
cis-india.org/papers/ebola-a-big-data-disaster

Are We Throwing Our Data Protection Regimes Under the Bus?
cis-india.org/internet-governance/blog/are-we-throwing-our-data-protection-regimes-under-the-bus

A Critique of Consent in Information Privacy
cis-india.org/internet-governance/blog/a-critique-of-consent-in-information-privacy

A Case for Greater Privacy Paternalism?
cis-india.org/internet-governance/blog/a-case-for-greater-privacy-paternalism

New Approaches to Information Privacy – Revisiting the Purpose Limitation Principle
cis-india.org/internet-governance/blog/digital-policy-portal-july-13-2016-new-approaches-to-information-privacy-revisiting-the-purpose-limitation-principle

About the Authors

AMBER SINHA

Amber works on issues surrounding privacy, big data, and cyber security. He is interested in the impact of emerging technologies like artificial intelligence and learning algorithms on existing legal frameworks, and how they need to evolve in response. Amber studied humanities and law at National Law School of India University, Bangalore.

ELONNAI HICKOK

Elonnai is a Director - Internet Governance at the Centre for Internet and Society, Bangalore. Elonnai has graduated from the University of Toronto where she studied international development and political science. Elonnai leads the privacy, surveillance, and big data work at the Centre and has also written extensively on issues pertaining to intermediary liability, digital rights, identity, cyber security and DNA profiling. Elonnai has also worked as a consultant with the Open Society Foundations and with the Ranking Digital Rights project.

ROHAN GEORGE

Rohan is penultimate year law student at the LSE. He also volunteers at Privacy international. He is interested in the intersection of big data technologies, privacy and how innovative big data solutions can be facilitated by creative approaches to privacy and data protection.

SCOTT MASON

Scott is a postgraduate researcher at Keele University (UK). His research examines the democratic legitimacy of multi-stakeholder internet governance organizations. He also regularly writes on a number of digital rights issues including privacy, access to knowledge, freedom of speech and censorship.

SEAN MARTIN MCDONALD

Sean Martin McDonald is the CEO of FrontlineSMS's social enterprise, the Social Impact Lab, and the founder of the FrontlineSMS: Legal project. Sean joined FrontlineSMS in 2010 and leads the day-to-day operation, strategic direction, and project-driven application of FrontlineSMS. Sean is an advisor to the Clinton Global Initiative, TechChange, Digital Democracy, and an affiliate expert with UNDP. He sits on the Board of Directors of International Peace Park Expeditions Foundation, the SIMLab CIC, and the SIMLab Foundation. Sean is a lawyer, licensed in New York, holding a J.D. and an M.A. in International Peace and Conflict Resolution from American University. He earned his bachelor's degree in Magazine Journalism, Government and Spanish at the University of Maryland.

VANYA RAKESH

Vanya graduated in Law from ILNU, Ahmedabad. She is working as a Programme Officer on Big Data implications in the Global South and Privacy with the Information Policy team. She has been trained at some of the top legal firms in India and has an inclination towards Internet Law. She is based out of Bangalore.

VIPUL KHARBANDA

Vipul Kharbanda is a consultant with the Center for Internet and Society, Bangalore. After finishing his BA.LLB.(Hons.) from National Law School of India University in Bangalore, he worked for India's largest corporate law firm for two and a half years in their Mumbai office for two years working primarily on the financing of various infrastructure projects such as Power Plants, Roads, Airports, etc. Since quitting his corporate law job, Vipul has been working as the Associate Editor in a legal publishing house which has been publishing legal books and journals for the last 90 years in India. He has also been involved with the Center for Internet and Society as a Consultant working primarily on issues related to privacy and surveillance.