DATA PROTECTION IN AN ERA OF BIG DATA: THE CHALLENGES POSED BY BIG PERSONAL DATA

MOIRA PATERSON* AND MAEVE McDONAGH**

Big Data involves analysis based on artificial intelligence and machine learning to mine vast troves of personal data to find correlations which are used to inform decisions that affect individuals. This raises privacy issues, as well as broader issues of lack of due process, discrimination and consumer protection. This article analyses the Privacy Act 1988 *(Cth) (*'Privacy Act*') and identifies a number of limitations in its capacity to address the issues posed by Big Personal Data. It then discusses three possible sources of solutions for these issues: the new* General Data Protection Regulation*, which commenced operation in the European Union in May 2018; relevant recommendations in the Productivity Commission's report on Data Availability and Use; and the proposed new criminal offences for re-identification of de-identified government data in the Privacy Amendment (Re-identification Offence) Bill 2016 (Cth). It also considers the extent to which other laws may have a role to play in addressing the gaps identified in the* Privacy Act.

I INTRODUCTION

Big Data represents a new frontier in the way in which information is processed and used to inform decision-making. At its core it involves the use of analytical tools based on artificial intelligence and machine learning to mine the vast data troves being gathered and accumulated at ever increasing rates. Its objective is to find 'small patterns'[1] or correlations that reveal new insights or truths.

While the epistemological claims for Big Data (in particular, the notion that 'the volume of data, accompanied by techniques that can reveal their inherent truth, enables data to speak for themselves free of theory')[2] remain open to dispute,[3] it is increasingly being used in relation to personal data to inform decisions that impact on many facets of people's lives. This use raises important privacy issues as well as broader issues of lack of due process, discrimination and consumer

* Professor, Faculty of Law, Monash University.

** Professor, Law School, University College Cork, Cork, Ireland.

[1] Luciano Floridi, 'Big Data and Their Epistemological Challenge' (2012) 25 *Philosophy & Technology* 435, 436.

[2] Rob Kitchin, 'Big Data, New Epistemologies and Paradigm Shifts' (2014) 1(1) *Big Data & Society* 1, 3.

[3] See, eg, ibid 3–5; Lawrence Busch, 'A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large-Scale Data Sets' (2014) 8 *International Journal of Communication* 1727; Kate Crawford, Kate Miltner and Mary L Gray, 'Critiquing Big Data: Politics, Ethics, Epistemology: Special Section Introduction' (2014) 8 *International Journal of Communication* 1663, 1668–70.

protection. This is problematic because it challenges the paradigms that form the basis for the key regimes that are currently used to protect the privacy of personal data — ie data protection laws.

The main data protection law in Australia is the *Privacy Act 1988* (Cth) ('*Privacy Act*'), which regulates the Commonwealth public sector and the private sectors. This is supplemented by various state and territory laws which regulate information handling by state and territory public sector bodies based on broadly similar sets of privacy principles.[4] Our focus is on the federal law and its regulation of Big Data outside of the specialist contexts of law enforcement and national security.[5]

The article identifies a number of key limitations in its capacity to address the issues posed by Big Personal Data. It then discusses three possible sources of solutions for these issues. First it considers the new *General Data Protection Regulation*,[6] which commenced operation in the European Union in May 2018, with specific reference to key features which offer potential solutions for Australia. It then considers relevant recommendations in the Productivity Commission's report on Data Availability and Use[7] and the proposed new criminal offences for re-identification of de-identified government data in the Privacy Amendment (Re-identification Offence) Bill 2016 (Cth). Finally, it considers the extent to which other laws may have a role to play in addressing the identified gaps in the *Privacy Act*.

II WHAT IS BIG PERSONAL DATA AND HOW IS IT USED?

In order to understand the challenges posed by Big Personal Data it is important to understand what is new and different about it. While the concept of Big Data is now reasonably familiar, it is not necessarily well understood and has been described as a 'generalized, imprecise term'.[8] Big Data is most commonly defined with reference to its key common characteristics, which are frequently described as 'volume, velocity, and variety'.[9] It is important to understand that the word

[4] *Information Privacy Act 2014* (ACT); *Privacy and Personal Information Protection Act 1998* (NSW); *Information Act 2002* (NT); *Information Privacy Act 2009* (Qld); *Personal Information Protection Act 2004* (Tas); *Privacy and Data Protection Act 2014* (Vic). These are supplemented in respect of health information by the *Health Records (Privacy and Access) Act 1997* (ACT); *Health Records and Information Privacy Act 2002* (NSW); *Health Records Act 2001* (Vic).

[5] We have excluded the latter due to the complexity of its regulation.

[6] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)* [2016] OJ L 119/1 ('*GDPR*').

[7] Productivity Commission, 'Data Availability and Use' (Inquiry Report No 82, Productivity Commission, 31 March 2017) ('*Data Availability and Use: Inquiry Report*').

[8] Kate Crawford and Jason Schultz, 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms' (2014) 55 *Boston College Law Review* 93, 96.

[9] Doug Laney, '3D Data Management: Controlling Data Volume, Velocity, and Variety' on *Gartner Blog Network* (6 February 2001) <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. See also the longer list provided in Kitchin, above n 2, 1–2.

'big' is not simply reflective of the fact that it involves large accumulations of data; what is significant is that it also involves a marked change in the ways in which data is analysed and used.[10]

Big Personal Data involves the application of speedy and sophisticated data analysis to huge data sets, including data about individuals and groups of individuals, gathered from a wide range of sources. It utilises analytical tools, commonly referred to as Big Data analytics ('BDA'), which commonly utilise artificial intelligence ('AI') — a process that analyses data to 'model some aspect of the world' and to draw inferences 'to predict and anticipate possible future events'.[11] As explained by the UK Information Commissioner, a significant feature of AI programs is that 'they learn from the data in order to respond intelligently to new data and adapt their outputs accordingly'.[12] A feature of AI that is used in BDA is that it uses complex mathematical algorithms both to process data and to make decisions based on that data. These algorithms are generally non-transparent, creating what has been described as 'a "black box" effect'.[13]

The BDA process seeks out correlations to offer new insights or make predictions about individuals and their behaviours, including insights and predictions which were unexpected prior to the analysis.[14] These are then used to facilitate a range of decisions and activities relating to individuals and groups of individuals.[15] It focuses on what correlations can be extracted from the data, rather than on their causation.[16]

The sources of data for the conduct of Big Personal Data analytics are many and varied. They include the large body of data that has become available as a by-product of human activities, as they 'have become mediated by digital services and devices'.[17] This results in a large body of so-called 'footprint' data: ie 'information

[10] While there is a view that the inclusion of enhanced analytical tools in the definition of Big Data creates definitional weakness, we take the view that they are integral to current initiatives and should therefore be regarded as a fundamental feature of what we describe as Big Personal Data: see further Pompeu Casanovas et al, 'Regulation of Big Data: Perspectives on Strategy, Policy, Law and Privacy' (2017) 7 *Health and Technology* 335, 336.

[11] Government Office for Science (UK), 'Artificial Intelligence: Opportunities and Implications for the Future of Decision Making' (Report, 9 November 2016) 5.

[12] Information Commissioner's Office (UK), 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (Discussion Paper, 4 September 2017) 7 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.

[13] Ibid 10. See also Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2015).

[14] Kitchin, above n 2, 4. See, eg, those between liking 'curly fries' on Facebook and being intelligent: Jennifer Golbeck, *Your Social Media 'Likes' Expose More than You Think* (October 2013) TED <https://www.ted.com/talks/jennifer_golbeck_the_curly_fry_conundrum_why_social_media_likes_say_more_than_you_might_think>; and between commuting to work by train and happiness: Office for National Statistics (UK), 'Commuting and Personal Well-Being, 2014' (Article, 12 February 2014) 16 <http://www.ons.gov.uk/ons/dcp171766_351954.pdf>.

[15] Ian Kerr and Jessica Earle, 'Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy' (2013) 66 *Stanford Law Review Online* 65.

[16] Andrej Zwitter, 'Big Data Ethics' (2014) 1(2) *Big Data & Society* 1, 2.

[17] Renaud Lambiotte and Michal Kosinski, 'Tracking the Digital Footprints of Personality' (2014) 102 *Proceedings of the IEEE* 1934, 1934.

that is given off by actions humans are already taking'.[18] We now also increasingly generate data as we make use of objects forming part of the 'Internet of Things'.[19] Take, for example, wearable objects such as 'Fitbit' devices and smart watches. These not only shed light on their wearers' 'physiological characteristics but … are also able to reconstruct the world around them by way of location coordinates, current speed travelled and direction, rich high-resolution photographs, and even in some cases audio capture'.[20] As summarised by Greenwood et al, these 'digital breadcrumbs we leave behind are clues to who we are, what we do, and what we want. This makes personal data — data about individuals — immensely valuable, both for public good and for private companies.'[21]

Another rich source is the social data on the internet, including the vast array of data publicly available on 'forums, social networks, review sites, newsgroups, and blogs'.[22] This adds an important dimension to Big Personal Data because of the role of the internet in enabling people.

In summary, the data collected now sheds light on most aspects of individuals' lives, including their purchases, hobbies, likes and dislikes, state of health and fitness, their social and family networks, their political views and their use of electronic devices and domestic appliances. It also sheds light on their interactions and patterns of movement across physical and networked spaces and even on their personalities. The mining of these large troves of data to seek out new correlations creates many potential uses for Big Personal Data, including uses which were not anticipated when the data was analysed. One such use involves finding novel solutions to problems; a correlation may suggest a possible solution to a specific problem. Correlations may also suggest ways to target activities that are designed to influence human behaviour, to inform and shape decisions relating to individuals, or to better direct activities designed to pre-empt specific risks or threats.

[18] Jim Thatcher, 'Living on Fumes: Digital Footprints, Data Fumes, and the Limitations of Spatial Big Data' (2014) 8 *International Journal of Communication* 1765, 1769.

[19] 'Simply put, this is the concept of basically connecting any device with an on and off switch to the Internet (and/or to each other). This includes everything from cellphones, coffee makers, washing machines, headphones, lamps, wearable devices and almost anything else you can think of.': Jacob Morgan, 'A Simple Explanation of "The Internet Of Things"', *Forbes* (online), 13 May 2014 <https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#2ae232d01d09>.

[20] M G Michael, Katina Michael and Christine Perakslis, 'Uberveillance and the Internet of Things and People' (Paper presented at the 2014 International Conference on Contemporary Computing and Informatics (IC^{3}I), Mysore, India, 27–29 November 2014) 1384.

[21] Daniel Greenwood et al, 'The New Deal on Data: A Framework for Institutional Controls' in Julia Lane et al (eds), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (Cambridge University Press, 2014) 192, 193.

[22] Leah Betancourt, *How Companies Are Using Your Social Media Data* (2 March 2010) Mashable <http://mashable.com/2010/03/02/data-mining-social-media/#7XUnW8SqWEqo>.

III WHAT ARE ITS POTENTIAL BENEFITS?

The potential benefits of Big Personal Data follow logically from its potential uses.

From a commercial perspective, data is an extremely valuable asset and a major 'source of value creation'.[23] The retail sector, in particular, has embraced Big Personal Data analysis as a means of facilitating the making of predictions as to what and when people are likely to buy.[24] In the United States, for example, Wal-Mart uses 'sales, pricing, and economic data, combined with demographic and weather data, to fine-tune merchandising … and anticipate appropriate timing of store sales'.[25] Big Personal Data is also used by retailers to adjust the prices of their goods based on demand and inventory: the US retail chain, Stage Stores, uses Big Personal Data for what is known as 'markdown optimization, which tells merchants the best time to cut the price of a particular item in a particular store'.[26] A further use of Big Personal Data by retailers is to assist them in adjusting their onsite advertising based on demand. For example, the billboards at drive-in restaurants advertise cheaper 'products that can be served up quickly' when queues are long, and 'higher-margin items that take longer to prepare' when queues are short.[27] More broadly, Big Data is used to evaluate credit risks[28] and to predict insurance outcomes.[29] The emerging field of 'people analytics' (otherwise known as 'talent analytics') also enables 'firms [to] determine which candidates to hire, how to help workers improve job performance, and how to predict when an employee might quit or should be fired'.[30]

From the perspective of public health and safety, Big Data can lead to the making of predictions on matters such as the spread of disease or the occurrence of natural disasters.[31] It can aid in the diagnosis of disease and the identification of adverse

[23] Ira S Rubinstein, 'Big Data: The End of Privacy or a New Beginning?' (2013) 3 *International Data Privacy Law* 74, 76.

[24] Omer Tene and Jules Polonetsky, 'Privacy in the Age of Big Data: A Time for Big Decisions' (2012) 64 *Stanford Law Review Online* 63, 65.

[25] Jeffrey F Rayport, 'Use Big Data to Predict Your Customers' Behaviors', *Harvard Business Review* (online), 5 September 2012 <https://hbr.org/2012/09/use-big-data-to-predict-your-c>.

[26] Teresa Meek, 'Big Data in Retail: How to Win with Predictive Analytics', *Forbes* (online), 18 February 2015 <http://www.forbes.com/sites/netapp/2015/02/18/big-data-in-retail/>.

[27] Nicole Laskowski, *Ten Big Data Case Studies in a Nutshell* (28 October 2013) SearchCIO <http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell>.

[28] Florentin Butaru et al, 'Risk and Risk Management in the Credit Card Industry' (Working Paper No 21305, National Bureau of Economic Research, June 2015) 24.

[29] Rick Swedloff, 'Risk Classification's Big Data (R)evolution' (2014) 21 *Connecticut Insurance Law Journal* 339, 341.

[30] Matthew T Bodie et al, 'The Law and Policy of People Analytics' (2017) 88 *University of Colorado Law Review* 961, 963; Mark Burdon and Paul Harpur, 'Re-Conceptualising Privacy and Discrimination in an Age of Talent Analytics' (2014) 37 *University of New South Wales Law Journal* 679, 681.

[31] Jeremy Ginsberg et al, 'Detecting Influenza Epidemics Using Search Engine Query Data' (2009) 457 *Nature* 1012; Henry Pearce, 'Online Data Transactions, Consent, and Big Data: Technological Solutions to Technological Problems?' (2015) 21 *Computer and Telecommunications Law Review* 149, 149.

side-effects of drugs.[32] It can predict 'energy demands to optimize renewable sources'[33] and can help to predict where traffic accidents and jams are likely to occur.[34] Big Data can also help to identify security and law enforcement risks.[35] From an individual perspective, Big Personal Data facilitates the customisation of services and products for individual consumers. Moreover, it can empower consumers by identifying the optimum time to purchase goods or services such as airline tickets.[36] More pragmatically, Big Data can help to support the development of online content without the levying of charges on consumers of such content.[37]

IV WHAT ARE ITS POTENTIAL HARMS?

The potential harms arising from Big Personal Data fall into two groups. The first is common to all Big Data and relates to its potential to mislead, either because of flaws in the data, the analytical tools applied to it, or because patterns identified are not true correlations.[38] As described by boyd and Crawford, 'Big Data enables the practice of apophenia: seeing patterns where none actually exist, simply because enormous quantities of data can offer connections that radiate in all directions'.[39] The second, which is the focus of this article, relates to its potential to undermine privacy. This is a harm in itself, but it also results in additional consequential harms.

Protecting the privacy of personal information is important because of the key role of privacy in protecting core values which underlie many other human rights. As explained by the Australian Law Reform Commission in its first report on privacy, the core feature shared by privacy and other core human rights is that it expresses 'the claim that each individual has [a right] to be treated as an autonomous human person, not just as an object or as a statistic'.[40]

[32] Tene and Polonetsky, above n 24, 64.

[33] Ibid.

[34] Ibid 64–5.

[35] Elizabeth E Joh, 'The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing' (2016) 10 *Harvard Law & Policy Review* 15, 16–17.

[36] Rubinstein, above n 23, 81–2; Thomas M Lenard and Paul H Rubin, 'Big Data, Privacy and the Familiar Solutions' (2015) 11 *Journal of Law, Economics & Policy* 1, 9.

[37] Alessandro Acquisti, Curtis Taylor and Liad Wagman, 'The Economics of Privacy' (2016) 54 *Journal of Economic Literature* 442, 454.

[38] See Nassim N Taleb, 'Beware the Big Errors of "Big Data"', *Wired* (online), 2 August 2013 <https://www.wired.com/2013/02/big-data-means-big-errors-people/>; danah boyd and Kate Crawford, 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon' (2012) 15 *Information, Communication & Society* 662, 667–70.

[39] boyd and Crawford, above n 38, 668. The authors refer to an example where it was 'demonstrated that data mining techniques could show a strong but spurious correlation between the changes in the S&P 500 stock index and butter production in Bangladesh': at 668, citing David J Leinweber, 'Stupid Data Miner Tricks: Overfitting the S&P 500' (2007) 16(1) *Journal of Investing* 15.

[40] Law Reform Commission, *Privacy*, Report No 22 (1983) vol 1, 13.

Big Personal Data is harmful to privacy because it removes the ability of individuals to exercise control over their own individual data, thereby undermining their autonomy (ie 'living and ordering a life of one's own choosing').[41] The concept of autonomy is central to liberal theory. Individual autonomy has been described by Christman as 'an idea that is generally understood to refer to the capacity to be one's own person, to live one's life according to reasons and motives that are taken as one's own and not the product of manipulative or distorting external forces'.[42] Except to the extent that it is based exclusively on analysis of data collected and used with the informed consent of the individuals concerned, Big Personal Data undermines the autonomy of data subjects in the processing of their data; it also facilitates activities and actions that further undermine autonomy by subjecting their decision-making to manipulation.

Autonomy is closely related to dignity. As stated by Feldman, 'one needs to take oneself and others seriously as moral agents' to be able to 'develop and exercise a capacity for self-determination'.[43] Respecting human dignity requires treating individuals as persons rather than things:

> Dignity is that which resists exchange; it is a thing that cannot be replaced by an equivalent. That is, there is a quality that renders human being that which cannot be commodified; to commodify this essential element is to strip the subject of his or her humanity, for dignity is one of humanity's essential qualities.[44]

Moreham further explains that '[i]t is this entitlement to respect, to be treated as an "end" and not simply as a "means", that many theorists argue underpins the privacy interest'.[45] Big Personal Data undermines human dignity by disregarding information subjects' choices as to how their personal information is used and their feelings concerning the ways in which their information is processed and used. More fundamentally, it undermines human dignity by treating individuals as objects for analysis and facilitating decision-making, which further objectifies them.

The consequential harms that flow from the privacy-invasiveness of Big Personal Data are even more far-reaching. It may expose individuals to decision-making based on possibilities, rather than probabilities or certainties, and, where the analysis results in flawed conclusions, to decision-making that is inherently faulty. As explained by boyd and Crawford, '[l]arge data sets from Internet sources are often unreliable, prone to outages and losses, and these errors and gaps are

[41] New Zealand Law Commission, *A Conceptual Approach to Privacy*, Miscellaneous Paper No 19 (2007) 5 [4.2].

[42] John Christman, *Autonomy in Moral and Political Philosophy* (26 January 2018) Stanford Encyclopedia of Philosophy <https://plato.stanford.edu/archives/spr2018/entries/autonomy-moral/>.

[43] David Feldman, 'Secrecy, Dignity, or Autonomy? Views of Privacy as a Civil Liberty' (1994) 47 *Current Legal Problems* 41, 54.

[44] Katie Foord, 'Defining Privacy' (Occasional Paper, Victorian Law Reform Commission, 1 January 2002) 19 <http://www.lawreform.vic.gov.au/sites/default/files/Defining_Privacy_Occasional_Paper.pdf>.

[45] N A Moreham, 'Why is Privacy Important? Privacy, Dignity and Development of the New Zealand Breach of Privacy Tort' in Jeremy Finn and Stephen Todd (eds), *Law, Liberty, Legislation: Essays in Honour of John Burrows QC* (LexisNexis, 2008) 231, 236.

magnified when multiple data sets are used together.'[46] Automated decision-making based on possibilities raises important due process issues; it may create outcomes that are unfair and also denies persons affected by it respect for their dignity as individuals.

A further issue of concern is that the social richness of Big Personal Data allows inferences about matters such as people's personalities,[47] and can assist in identifying personal weaknesses which can potentially be exploited to manipulate their behaviour. One example relates to its use in political campaigning, typified by allegations concerning the use of analytics to influence the outcome of the Brexit referendum and US presidential elections in 2016.[48] O'Neil describes the use of Big Data profiling in this context as the 'ultimate example of asymmetric information', permitting politicians to manipulate votes and donations.[49] Another example identified by Calo relates to its potential use to encourage irrational behaviour on the part of consumers by exploiting information about their weaknesses and emotions.[50]

Decision-making based on Big Personal Data also exposes individuals and groups to differential treatment (for example, price discrimination based on differential discounts).[51] This involves discrimination in the sense that it allows decision-makers to draw fine-grained distinctions between individuals which are then used as a basis for differential treatment. While such practices are commonplace in some sectors, for example, the insurance sector, Big Personal Data permits their more widespread use in relation to information which has not previously been available. This raises important questions as to whether there are 'specific differences' additional to those currently protected by anti-discrimination laws which should not be ignored.[52]

A further consequential harm arises when pre-emptive predictions based on Big Personal Data 'are intentionally used to diminish a person's range of future options',[53] thereby undermining their civil liberties. One example cited by Kerr and Earle involves the use of pre-emptive predictions to compile a no-fly list to address the risk of terrorist activity on planes.[54] This is problematic for the reason that '[i]n addition to curtailing liberty, a no-fly list that employs predictive

[46] boyd and Crawford, above n 38, 668.

[47] Lambiotte and Kosinski, above n 17.

[48] Jamie Doward and Alice Gibbs, 'Did Cambridge Analytica Influence the Brexit Vote and the US Election?' *The Guardian* (online), 5 March 2017 <https://www.theguardian.com/politics/2017/mar/04/nigel-oakes-cambridge-analytica-what-role-brexit-trump>.

[49] Cathy O'Neil, 'Big-Data Algorithms Are Manipulating Us All', *Wired* (online), 18 October 2016 <https://www.wired.com/2016/10/big-data-algorithms-manipulating-us/>.

[50] See Ryan Calo, 'Digital Market Manipulation' (2014) 82 *George Washington Law Review* 995, 1010.

[51] See Council of Economic Advisers, 'Big Data and Differential Pricing' (Report, Executive Office of the President of the United States, February 2015) <https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf>.

[52] See Tal Z Zarsky, 'Understanding Discrimination in the Scored Society' (2014) 89 *Washington Law Review* 1375, 1382, citing Frederick Schauer, *Profiles, Probabilities, and Stereotypes* (Harvard University Press, 2006) 215.

[53] Kerr and Earle, above n 15, 67.

[54] Ibid.

algorithms preempts the need for any evidence or constitutional safeguards. Prediction simply replaces the need for proof.'[55] Similar issues arise in relation to the use of Big Personal Data to inform 'predictive policing' strategies whereby police 'select what streets, groups, and individuals to subject to extra scrutiny'.[56] Such uses may erode civil liberties, such as the presumption of innocence.

Another problem is that the large-scale accumulation of personal data required to facilitate Big Personal Data activities also has the potential to result in security breaches which may expose individuals to identity theft and fraud.

V THE LIMITATIONS OF THE *PRIVACY ACT* IN ADDRESSING THE CHALLENGES POSED BY BIG PERSONAL DATA

The *Privacy Act* is designed to regulate the handling of personal information via a set of information privacy principles which regulate various aspects of information handling, including principles that impose limitations on the collection, use and disclosure of personal information, principles designed to make information handling more transparent, and principles that require organisations to keep personal information secure. The limitations on collection, use and disclosure are more extensive in the case of those categories of personal information which also qualify as 'sensitive information'.[57]

A *Private Sector Coverage*

The Act's coverage of private sector organisations is limited by a number of exceptions. One which significantly reduces its coverage is an exception for small business operators[58] (ie businesses with a gross annual turnover of $3 million or less).[59] While that exception is itself subject to a number of important exceptions, including an exception for businesses that trade in personal information,[60] it has the consequence that smaller businesses which rely on Big Personal Data are not subject to regulation.

Other exceptions of relevance to this topic include: an exception for political parties, which is significant given the alleged role of Big Personal Data in the

[55] Ibid 69.

[56] Ananda Mitra, 'Narbs: A Narrative Approach to the Use of Big Data' (2014) 38 *Annals of the International Communication Association* 369, 381, quoting Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt, 2013) 158.

[57] 'Sensitive information' is defined in the *Privacy Act* s 6(1).

[58] Ibid s 6C (definition of 'organisation').

[59] Ibid ss 6D–6DA.

[60] Ibid s 6D(4)(c).

Trump election and Brexit referendum;[61] and an exception for employee data which limits its ability to regulate uses of Big Personal Data by employers.[62]

B *The Definition of Personal Data*

An important feature of the *Privacy Act* is that it is limited in its scope to 'personal information', which is defined as 'information or an opinion about an identified individual, or an individual who is reasonably identifiable: (a) whether the information or opinion is true or not; and (b) whether the information or opinion is recorded in a material form or not'.[63] This limits the ability of the Act to address Big Personal Data in two specific ways.

The first issue relates to its individual focus, which is an accepted feature of most privacy laws. This is problematic because of the potential for decisions and activities based on Big Personal Data to result in group level harms, including discrimination, in cases where information is collected on a group basis and does not qualify for protection as personal information.[64]

This issue of group privacy has been explored in the context of assessment of the usefulness of anonymisation techniques to protect privacy. Zwitter makes the point that

> [d]e-individualization (ie removing elements that allow data to be connected to one specific person) is … just one aspect of anonymization. Location, gender, age, and other information relevant for the belongingness to a group and thus valuable for statistical analysis relate to the issue of group privacy.[65]

It follows that '[a]nonymization of data is, thus, a matter of degree [in terms] of how many and which group attributes remain in the data set … [and that] groups are always becoming more transparent', despite the anonymisation of data.[66]

The specific harm that can result from invasion of group privacy has been explored most extensively in the context of bioethics, where it has been explained as follows:

> The harm caused by associating an increased risk of a stigmatizing condition (eg, mental illness) with a particular group, especially a minority or socially vulnerable group, attaches to each member of the group regardless of whether that person's health record or biological specimen was used in the research. Even

[61] See, eg, Roberto J González, 'Hacking the Citizenry? Personality Profiling, "Big Data" and the Election of Donald Trump' (2017) 33(3) *Anthropology Today* 9; William Davies, 'How Statistics Lost Their Power — and Why We Should Fear What Comes Next', *The Guardian* (online), 19 January 2017 <https://www.theguardian.com/politics/2017/jan/19/crisis-of-statistics-big-data-democracy>.

[62] For a detailed discussion of the employee records exception in *Privacy Act* s 7B(3), see Australian Law Reform Commission, *For Your Information: Australian Privacy Law and Practice*, Report No 108 (2008) vol 2, 1363–411.

[63] *Privacy Act* s 6 (definition of 'personal information').

[64] Brent Mittelstadt, 'From Individual to Group Privacy in Big Data Analytics' (2017) 30 *Philosophy & Technology* 475.

[65] Zwitter, above n 16, 4.

[66] Ibid.

> if certain members of the group provided informed consent for research with their samples, any resulting stigma would be shared with nonparticipating members of the group.[67]

The second issue is that the definition of 'personal information' is limited to information about individuals who are identified or potentially identifiable and does not therefore apply to information that has been de-identified. This is problematic because Big Personal Data challenges the validity of the assumption which underlies this limitation, ie that anonymisation of data removes the concerns that privacy protection is designed to achieve.

First, its associated computational technology increasingly permits the re-identification of data that has been de-identified. As noted by Hardy and Maurushat, 'new techniques for de-identifying data have been met with equally innovative attempts at re-identification'.[68]

The limitations of anonymisation in the face of re-identification attacks have been demonstrated in a number of studies,[69] although there is continuing work on approaches designed to limit potential threats by 'abstracting or perturbing data'.[70] Based on a review of the literature on this topic, the UN Special Rapporteur has suggested that while '[s]imple kinds of data, such as aggregate statistics, are amenable to genuinely privacy-preserving treatment such as differential privacy',[71] '[h]igh-dimensional unit-record level data cannot be securely de-identified without substantially reducing its utility'.[72] This raises the question of whether it is safe to rely on de-identification as a means of protecting personal privacy and, therefore, whether to remove de-identified data altogether from the scope of data protection regimes.[73]

Second, even where data is processed in ways which do not reveal an individual's real-life identity, the richness of the data which can be associated with an

[67] Mark A Rothstein, 'Is Deidentification Sufficient to Protect Health Privacy in Research?' (2010) 10(9) *American Journal of Bioethics* 3, 6.

[68] Keiran Hardy and Alana Maurushat, 'Opening up Government Data for Big Data Analysis and Public Benefit' (2017) 33 *Computer Law & Security Review* 30, 32, citing Jules Polonetsky, Omer Tene and Kelsey Finch, 'Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification' (2016) 56 *Santa Clara Law Review* 593, 594.

[69] See, eg, Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 *UCLA Law Review* 1701; Arvind Narayanan and Vitaly Shmatikov, 'Robust De-Anonymization of Large Sparse Datasets' (Paper presented at the 2008 IEEE Symposium on Security and Privacy, Oakland, California, USA, 18–21 May 2018) 111–25.

[70] Solon Barocas and Helen Nissenbaum, 'Big Data's End Run around Anonymity and Consent' in Julia Lane et al (eds), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (Cambridge University Press, 2014) 44, 50. These approaches include *k*-anonymity (see, eg, Latanya Sweeney, '*k*-Anonymity: A Model for Protecting Privacy' (2002) 10 *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 557) and differential privacy (see, eg, Cynthia Dwork and Aaron Roth, *The Algorithmic Foundations of Differential Privacy* (Now Publishers, 2014)).

[71] Joseph A Cannataci, *Right to Privacy*, 72nd sess, Agenda Item 72(b), UN Doc A/72/540 (19 October 2017) 18 [96].

[72] Ibid [97]. See also Joseph A Cannataci, *Supporting Documents* (19 October 2017) Office of the United Nations High Commissioner on Human Rights, 1–6 <www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx>.

[73] See also Pompeu Casanovas et al, 'Regulation of Big Data: Perspectives on Strategy, Policy, Law and Privacy' (2017) 7 *Health and Technology* 335, 339.

individual permits a multiplicity of inferences which impact on how they are treated, even when the decision-maker is unaware of who they are.[74] This is illustrated by the following example:

> A website uses a formula to turn its users' email addresses into jumbled strings of numbers and letters. An advertiser does the same with its customer email lists. Both then send their jumbled lists to a third company that looks for matches. When two match, the website can show an ad targeted to a specific person, but no real email addresses changed hands.[75]

The issue here is that anonymity again offers no protection and still permits the harms to autonomy and dignity which form the basis for privacy protection.

A final issue relates the extent to which the definition of 'personal information' in the Act[76] applies to information that is linked to an individual and allows 'reach' of an individual (in the sense of making decisions or taking actions that affect him or her) but which is not clearly information about a person. This was considered recently in the context of litigation by technology journalist Ben Grubb relating to his privacy rights to access the personal metadata held by his mobile phone service provider, which was potentially subject to access by national security and law enforcement agencies. The Full Court of the Federal Court upheld the view expressed by the Administrative Appeals Tribunal that the words 'about an individual' in the definition of personal information raised a threshold question that needed to be addressed before it could be determined whether that individual is identified or identifiable.[77]

In obiter dicta, Kenny and Edelman JJ stated:

> The words 'about an individual' direct attention to the need for the individual to be a subject matter of the information or opinion. This requirement might not be difficult to satisfy. Information and opinions can have multiple subject matters. Further, on the assumption that the information refers to the totality of the information requested, then even if a single piece of information is not 'about an individual' it might be about the individual when combined with other information. However, in every case it is necessary to consider whether each item of personal information requested, individually or in combination with other items, is about an individual. This will require an evaluative conclusion, depending upon the facts of any individual case, just as a determination of whether the identity can *reasonably* be ascertained will require an evaluative conclusion.[78]

This approach is open to criticism because it removes data that is not currently 'about' an individual from the operation of the Act even where that information

[74] These are discussed in detail in Barocas and Nissenbaum, above n 70, 52–6.

[75] Ibid 53, quoting Jennifer Valentino-DeVries and Jeremy Singer-Vine, 'They Know What You're Shopping for' *The Wall Street Journal* (online), 7 December 2012 <https://www.wsj.com/articles/SB10001424127887324784404578143144132736214>.

[76] To receive protection under the Act information must be 'about' a specific individual: see *Privacy Act* s 6(1) (definition of 'personal information').

[77] *Privacy Commissioner v Telstra Corporation Ltd* (2017) 249 FCR 24, 35–7 [60]–[65].

[78] Ibid 36 [63] (emphasis in original).

is potentially capable of revealing information about that individual. This is significant in the context of Big Personal Data as it has the consequence that data sets and the information produced from them via Big Data analytics will remain unregulated until that data is applied to specific individuals.

C *The Limited Restrictions on Collection and Subsequent Uses and Disclosures of Personal Information*

A distinguishing feature of the collection, use and disclosure limitation principles is that they are not, in the main, based on a consent model.

In the case of personal information which is not also sensitive information, the main restrictions on collection are that the information must be collected only if 'the information is reasonably necessary for, or directly related to, one or more of' an APP entity's functions or activities,[79] and 'by lawful and fair means'.[80] Furthermore, once this information has been collected it can be used and disclosed consistently with the primary purpose for its collection without the need to obtain any further consent.[81]

The requirement that information collection must be reasonably necessary for, or directly related to, one or more of a collector's functions or activities is arguably inadequate in relation to the collection of personal information by organisations whose functions or activities are related to Big Personal Data. It does nothing to restrict the collection and use of information for profiling purposes.

Similarly, the requirement that collection must be fair and lawful poses minimal restrictions on common forms of non-consent based collections of personal data such as collections from publicly available sources. This is significant given the extent to which Big Personal Data relies on collection of personal information from social media sites.

Consent is generally required for the collection of sensitive information.[82] It is also required for uses and disclosures for secondary purposes unless: the individual would reasonably expect that use; and that use is related to the primary purpose (in the case of personal information that is not sensitive information) or directly related to the primary purpose (in the case of sensitive information).[83] This limitation on use of personal information for secondary purposes is based on the purpose specification principle, whereby data must be collected for specific, explicit and lawful purposes and must not be used for any other incompatible purpose.[84] The purpose specification principle creates difficulties in the context of Big Personal Data 'where personal data moves freely through a revolving door

[79] *Privacy Act* sch 1 cl 3.

[80] Ibid sch 1 cl 3.5.

[81] Ibid sch 1 cl 6.1.

[82] Ibid sch 1 cl 3.3.

[83] Ibid sch 1 cls 6.1–6.2.

[84] Bart Custers and Helena Uršič, 'Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection' (2016) 6 *International Data Privacy Law* 4, 8–9.

from public to private', public to public and private to private bodies, 'with little regard for the purposes for which it was originally collected',[85] and where data analytics are strongly dependent on the 'repurposing' and 'recontextualization'[86] of data which, in many instances, will amount to incompatible processing.

The data minimisation principle, another core data protection principle, holds that that data should be adequate, relevant and not excessive in relation to the purposes for which they are processed.[87] It too has been described as being 'inimical to the underlying thrust of [Big Data], which discovers new correlations by applying sophisticated analytic techniques to massive data collection, and seeks to do so free of any *ex ante* restrictions'.[88]

Furthermore, a problem with Big Personal Data is that it makes it difficult to expect individuals to provide consent at the time of collection because they may not know what they are consenting to, while bodies handling information may not know what to seek consent for.[89] Mantelero ascribes this problem to 'the "transformative" use of Big Data', which makes it 'often impossible to explain … all the possible uses of information at the time of its initial collection'.[90]

The complexity of data processing also renders reliance on consent problematic. Cate and Mayer-Schönberger point out that the processing of data has become very 'complicated as datasets are combined and data processors and users change',[91] rendering it difficult for data controllers to provide sufficient information to data subjects so as to obtain informed consent. Consent notices can also be overly complex.[92]

The issue of consent is complicated by the imbalance of power that can exist between the data subject and the data controller. This is the case in the context of relationships such as employer-employee and landlord-tenant but the issue is also present in the interactions between consumers and large companies and between citizens and state apparatus such as law enforcement and security authorities. Such imbalances are readily observed in 'the business terms and privacy policies

[85] Judith Rauhofer, 'Round and Round the Garden? Big Data, Small Government and the Balance of Power in the Information Age' (2015) 8(1) 경제규제와 법 [*Journal of Law & Economic Regulation*] 34, 43.

[86] Custers and Uršič, above n 84, 8–9.

[87] *GDPR* [2016] OJ L 119/1, art 5(1)(c).

[88] Rubinstein, above n 23, 78.

[89] See Daniel J Solove, 'Introduction: Privacy Self-Management and the Consent Dilemma' (2013) 126 *Harvard Law Review* 1880, 1902.

[90] Alessandro Mantelero, 'The Future of Consumer Data Protection in the EU: Re-Thinking the "Notice and Consent" Paradigm in the New Era of Predictive Analytics' (2014) 30 *Computer Law & Security Review* 643, 645, quoting Tene and Polonetsky, above n 24, 64.

[91] Fred H Cate and Viktor Mayer-Schönberger, 'Notice and Consent in a World of Big Data' (2013) 3 *International Data Privacy Law* 67, 67–8.

[92] Cate suggests that in the US in particular notices tend to be overly complex: Fred H Cate, 'The Limits of Notice and Choice' (2010) 8(2) *IEEE Security & Privacy* 59, 60.

of online providers [which] are normally drafted in favour of those providers and are not negotiable'.[93]

Another problem is that it may be said that there is no real choice when it comes to consenting to the use of one's data: this phenomenon is referred to by Mantelero as 'social lock-in'.[94] The only choice the user is offered most of the time is to 'take-it-or-leave-it'.[95]

The consent model can also be criticised as not working in the way it was intended to in that it does not protect us from 'our own bad, ignorant, unintentional, or unavoidable choices'.[96] Solove notes that the notice and consent model 'envisions an informed and rational person who makes appropriate decisions about whether to consent to various forms of collection, use, and disclosure of personal data' but there is a lack of evidence of 'people's actual ability to make such informed and rational decisions'.[97]

VI ASPECTS OF THE *GDPR* THAT PROVIDE ADDITIONAL SOLUTIONS IN RESPECT OF BIG PERSONAL DATA

The *GDPR* builds on features of the previous EU Data Protection Directive,[98] which offered stronger privacy protection in many respects than that available under the Australian *Privacy Act*. The *GDPR* contains a number of features which have the potential to improve the regulation of Big Personal Data. These apply with reference to data controllers (ie a person or body which 'determines the purposes and means of the processing of personal data')[99] and data processors (ie a 'person … or other body which processes personal data on behalf of the controller').[100]

A *Clearer Application to Online Identifiers*

The definition of 'personal data' in art 4(1) of the *GDPR* is similar to the definition of 'personal information' in the *Privacy Act* in that it is confined to individuals and focuses on identifiability. However, a key difference is that it is defined to mean

[93] Judith Rauhofer, 'One Step Forward, Two Steps Back? — Critical Observations on the Proposed Reform of the EU Data Protection Framework' (2013) 6(1) 경제규제와 법 [*Journal of Law & Economic Regulation*] 57, 76.

[94] Mantelero, above n 90, 645.

[95] Rauhofer, 'One Step Forward, Two Steps Back?', above n 93, 76.

[96] Cate, above n 92, 61.

[97] Solove, above n 89, 1883.

[98] *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] OJ L 281/31.

[99] *GDPR* [2016] OJ L 119/1, art 4(7).

[100] Ibid art 4(8).

‘any information *relating to* an identified or identifiable natural person’.[101] The requirement that the information relate to a natural person has been interpreted as meaning that ‘by reason of its content, purpose or effect, [it] is linked to a particular person’.[102] The *GDPR* definition also makes clear that an identifiable natural person includes ‘one who can be identified, directly or indirectly’. It has therefore been interpreted in the case of dynamic IP addresses as requiring an assessment of whether there is a reasonable likelihood of linkage with other databases which will result in identification, rather than requiring assessment of the addresses in isolation from other linking data.[103]

The *GDPR* definition of ‘personal data’ also differs from the *Privacy Act* definition of ‘personal information’ in that it specifically addresses the issue of online identifiers and states that an ‘identifiable natural person’ includes ‘one who can be identified, directly or indirectly, in particular by reference to an identifier … [including] an online identifier’.[104] In addition, recital 30 makes clear that online identifiers such as those ‘provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags’ will be personal data where used to create profiles of people and identify them.[105]

Also of significance is recital 26, which states that in order ‘[t]o determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly’.[106] The express reference to ‘singling out’ suggests that the processing of data that singles out but does not reveal an individual’s identity comes within the scope of European data protection law.[107]

[101] Ibid art 4(1) (emphasis added).

[102] *Peter Nowak v Data Protection Commissioner* (European Court of Justice, C-434/16, 20 December 2017) [35].

[103] See *Patrick Breyer v Bundesrepublik Deutschland* (European Court of Justice, C-582/14, 19 October 2016) [49].

[104] *GDPR* [2016] OJ L 119/1, art 4(1).

[105] This strengthens the approach taken by the ECJ in *Patrick Breyer v Bundesrepublik Deutschland* (European Court of Justice, C-582/14, 19 October 2016). For a useful discussion of this case see Frederik Zuiderveen Borgesius, ‘The *Breyer* Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition’ (2017) 3 *European Data Protection Law Review* 130.

[106] *GDPR* [2016] OJ L 119/1, recital 26.

[107] This is consistent with the view expressed by the Article 29 Data Protection Working Party in *Opinion 4/2007 on the Concept of Personal Data* (20 June 2007) European Commission, 14 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf>. See also Frederik J Zuiderveen Borgesius, ‘Singling Out People without Knowing Their Names: Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation’ (2016) 32 *Computer Law & Security Review* 256.

B *Higher Bars for Collection, Use and Disclosure Which Is Justifiable Based on the Interests of the Collector*

The *GDPR* follows a similar pattern to the Australian *Privacy Act* in that it distinguishes between ordinary personal data and sensitive personal data (referred to in the *GDPR* as 'special categories' of personal data).[108] In the case of the former it permits collection, use and disclosure based on the interests of the collector, whereas in the latter explicit consent is generally required.

Article 6 of the *GDPR*, which sets out the permitted legal justifications for 'processing' (a term which encompasses collection, use and disclosure),[109] permits processing of personal data (other than special categories of data) based on 'the legitimate interests pursued by the controller or by a third party'.[110] What is meant by a legitimate interest is explained in recital 47, which states that it can exist 'where there is a relevant and appropriate relationship between the data subject and the controller in situations such as where the data subject is a client or in the service of the controller'.[111] Whether a legitimate interest exists in a specific case requires 'careful assessment including whether a data subject can reasonably expect at the time and in the context of the collection of the personal data that processing for that purpose may take place'.[112] The recital also makes clear that data processing 'for direct marketing purposes may be regarded as carried out for a legitimate interest'.[113]

Significantly, however, art 6 also requires that such legitimate interests are not 'overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child'.[114] Furthermore, recital 69 provides that 'a data subject should, nevertheless, be entitled to object to the processing of any personal data relating to his or her particular situation' and that '[i]t should be for the controller to

[108] Ie 'personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation': *GDPR* [2016] OJ L 119/1, art 9(1).

[109] '[P]rocessing' is defined in ibid art 4(2) as:

> any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction …

[110] This legitimate interests ground does not apply to processing carried out by public authorities, which is covered by a separate ground which permits processing that 'is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller': ibid art 6(1)(e).

[111] Ibid recital 47.

[112] Ibid.

[113] Ibid.

[114] Ibid art 6(1)(f). See *Valsts policijas Rīgas reģiona pārvaldes Kārtības policijas pārvalde v Rīgas pašvaldības SIA 'Rīgas satiksme'* (European Court of Justice, C-13/16, 4 May 2017). Processing is also permissible under the grounds specified in ibid arts 6(1)(a)–(e).

demonstrate that its compelling legitimate interest overrides the interests or the fundamental rights and freedoms of the data subject.'[115]

The requirements for a legitimate interest on the part of the collector or processor and for balancing such an interest against the interests and fundamental rights of data subjects has the potential to provide a valuable brake on Big Personal Data activities that are harmful to data subjects. The reasonable expectations of data subjects based on their relationship with the controller are an important factor that affects the balance between the interests of the controller and the rights of the data subjects. It follows therefore that the legitimate interest of the controller may be overridden by the interests and fundamental rights of the data subject in cases 'where personal data are processed in circumstances where data subjects do not reasonably expect further processing'.[116]

In the case of 'special categories of data' (which are generally equivalent to 'sensitive data' in the Australian *Privacy Act*), these grounds are not applicable and, subject to a number of exceptions,[117] processing is permissible only where 'the data subject has given explicit consent to the processing … for one or more specified purposes'.[118] This approach is generally similar to that under the Australian *Privacy Act*.

Significantly, however, consent must be 'freely given, specific, informed and [an] unambiguous indication of the data subject's wishes'.[119] Recital 43 seeks to address the issue of imbalance of power between data subjects and controllers by stating inter alia that 'consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller, in particular where the controller is a public authority'.[120] Recital 43 further provides that:

> Consent is presumed not to be freely given if it does not allow separate consent to be given to different personal data processing operations despite it being appropriate in the individual case, or if the performance of a contract, including the provision of a service, is dependent on the consent despite such consent not being necessary for such performance.[121]

These requirements provide for stronger protection, but it is open to query how consent can operate effectively in the context of Big Personal Data given that it may be difficult for data collectors to predict in advance the purposes for which data will be used and for data subjects to give informed consent to proposed uses.

[115] *GDPR* [2016] OJ L 119/1, recital 69.

[116] Ibid recital 47.

[117] Ibid arts 9(2)(b)–(j).

[118] Ibid art 9(2)(a).

[119] Ibid art 4(11).

[120] Ibid recital 43.

[121] Ibid.

C Changes to the Key Data Protection Principles

1 *Purpose Limitation Principle*

The *GDPR* introduces a test for determining whether processing is compatible with the purpose for which data were initially collected.[122] While the adoption of this test may be helpful in terms of guiding data controllers as to the circumstances in which they need to revert to data subjects to seek consent to processing for new purposes, it does not address the underlying problem that Big Personal Data will inevitably fall foul of the purpose limitation principle. A further issue is that the purpose limitation principle in the *GDPR* states that the further processing of personal data for 'statistical purposes' shall 'not be considered to be incompatible with the initial purposes' of collection of the data concerned.[123] This exception has been described as 'enabling Big Data without explicitly abandoning the purpose limitation principle'.[124]

2 *Data Minimisation Principle*

The data minimisation principle has been somewhat strengthened by the *GDPR* to the extent that rather than requiring that data should be 'adequate, relevant and not excessive' in relation to the purposes for which the data is collected and/or further processed,[125] the *GDPR* requires that it be 'adequate, relevant and limited to what is necessary' in relation to such purposes.[126] This alteration does not however address the underlying issue that the data minimisation principle is incompatible with the activity which is at the core of Big Personal Data analytics, namely the collection of vast quantities of data for purposes that may not be anticipated at the time of such collection. The *GDPR* also provides for an exception to the data minimisation principle which permits the longer retention of personal data for 'statistical purposes'.[127] The introduction of this exception has prompted Mayer-Schönberger and Padova to suggest that '[a]s arguably most, if not all, of Big Data

[122] This test, which is found in art 6(4) and is derived from Article 29 Data Protection Working Party, *Opinion 03/2013 on Purpose Limitation* (2 April 2013) European Commission, 23–7 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf>, requires the controller to take account of the following:

> (a) any link between the purposes for which the personal data have been collected and the purposes of the intended further processing; (b) the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller; (c) the nature of the personal data, in particular whether special categories of personal data … or … personal data related to criminal convictions and offences are processed … (d) the possible consequences of the intended further processing for data subjects; (e) the existence of appropriate safeguards, which may include encryption or pseudonymisation.

[123] *GDPR* [2016] OJ L 119/1, art 5(1)(b).

[124] Viktor Mayer-Schönberger and Yann Padova, 'Regime Change? Enabling Big Data through Europe's New Data Protection Regulation' (2016) 17 *Columbia Science & Technology Law Review* 315, 326.

[125] *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] OJ L 281/31, art 6(1)(c).

[126] *GDPR* [2016] OJ L 119/1, art 5(1)(c).

[127] Ibid art 5(1)(e).

analysis is statistical in nature, the *GDPR* offers an explicit pathway for Big Data analyses to work with retained data'.[128]

D *Requirements for Privacy by Design and Default*

In the case of potentially high-risk processing activities (a term which arguably encompasses much of Big Personal Data), art 25 of the *GDPR* requires the controller to build in privacy by design both at the time of the determination of the means for processing and at the time of the processing itself. Specifically, the controller must:

> implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.[129]

What specifically is required must be determined having regard to 'the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing'.[130]

Insofar as the issue of identifiability is concerned, recital 26 states that:

> To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.[131]

It would seem to follow therefore that issues concerning the risk of re-identification, having regard to the current state of the art, must be taken into account before relying on anonymisation to avoid obligations under the *GDPR*.

The Irish Data Protection Commission has commented in relation to this issue that:

> It is not normally possible to quantify the likelihood of re-identification of individuals from anonymised data. However, thinking about the risks which are present will assist in assessing whether identification of data subjects from anonymised data is likely. An effective anonymisation technique will be able to prevent the singling out of individual data subjects, the linking of records

[128] Mayer-Schönberger and Padova, above n 124, 330.

[129] *GDPR* [2016] OJ L 119/1, art 25(1).

[130] Ibid.

[131] Ibid recital 26.

> or matching of data between data sets, and inference of any information about individuals from a data set.[132]

The obligation to implement privacy by design is supplemented by a further duty to build in privacy by default (ie to 'implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed').[133] Again, this offers a higher level of protection in relation to Big Personal Data, although it may present difficulties given the amorphous purpose of Big Data.

E *Requirement to Conduct Data Protection Impact Assessments*

The *GDPR* also contains an additional requirement for a controller to conduct a data protection impact assessment ('DPIA') '[w]here a type of processing … is likely to result in a high risk to the rights and freedoms of natural persons'.[134] This assessment must include at minimum:

> (a) a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;
>
> (b) an assessment of the necessity and proportionality of the processing operations in relation to the purposes;
>
> (c) an assessment of the risks to the rights and freedoms of data subjects … and
>
> (d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.[135]

While the *GDPR* does not exhaustively define the situations in which such assessments are required, it is significant that the list of examples provided includes automated processing for purposes of profiling and similar activities intended to evaluate personal aspects of data subjects.[136]

The DPIA process is open to criticism on the basis that it is narrow and focuses only on compliance with the *GDPR*, rather than broader privacy issues.[137]

[132] Data Protection Commission (Ireland), *Anonymisation and Pseudonymisation* <https://www.dataprotection.ie/docs/Anonymisation-and-pseudonymisation/g/1594.htm>.

[133] *GDPR* [2016] OJ L 119/1, art 25(2).

[134] Ibid art 35(1).

[135] Ibid art 35(7).

[136] Ibid art 35(3)(a).

[137] See Claudia Quelle, 'The Data Protection Impact Assessment, or: How the *General Data Protection Regulation* May Still Come to Foster Ethically Responsible Data Processing' (Paper, Tilburg Institute for Law, Technology, and Society, 25 November 2015) 2 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2695398>; Roger Clarke, 'The Distinction between a PIA and a Data Protection Impact Assessment (DPIA) under the EU *GDPR*' (Paper presented at the Computers, Privacy and Data Protection 10th International Conference, Brussels, 27 January 2017) <http://www.rogerclarke.com/DV/PIAvsDPIA.html>.

Furthermore, it embodies a risk-based approach to the protection of personal data, which is problematic in that it 'results in leaving data protection issues mainly to data controllers to decide, while data protection authorities see their supervisory role significantly weakened'.[138] However, it is significant that the DPIA process requires assessment of the risks to the rights and freedoms of data subjects. The Article 29 Working Party has stated that this requirement 'primarily concerns the right to privacy … [and] may also involve other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion'.[139]

F *Special Provisions Regulating Profiling Based on Automatic Processing*

The *GDPR* contains a number of additional obligations and restrictions in relation to 'profiling', which is defined as

> any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.[140]

It is important to note that these protections are not available where the processing involves some element of human intervention.[141]

First, art 13 requires that data subjects must be informed inter alia about 'the existence of [any] automated decision making [and] … profiling'.[142] They must also be provided with 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing' for them.

Second, art 21 confers a right to object to data processing in specified circumstances which must be 'explicitly brought to the attention of the data subject' at the time of the first communication with them.[143] It provides data subjects with a right to object at any time to processing of their data, based on grounds relating to

[138] Maria Eduarda Gonçalves, 'The EU Data Protection Reform and the Challenges of Big Data: Remaining Uncertainties and Ways Forward' (2017) 26 *Information & Communications Technology Law* 90, 114.

[139] Article 29 Data Protection Working Party, *Statement on the Role of a Risk-Based Approach in Data Protection Legal Frameworks* (30 May 2014) European Commission, 4 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf>.

[140] *GDPR* [2016] OJ L 119/1, art 4(4).

[141] Guidelines produced by the Article 29 Working Party advise to qualify as human involvement, the controller must ensure that any 'oversight of the decision is meaningful', rather than just a token gesture and that '[i]t should be carried out by someone who has the authority and competence to change the decision': Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679* (6 February 2018) European Commission, 21 <http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053>.

[142] *GDPR* [2016] OJ L 119/1, art 13(2)(f).

[143] Ibid art 21(4).

their 'particular situation'.[144] This obligation arises where the processing is based on legitimate interests of the collector (or on the ground that it is necessary in the public interest or in the exercise of official authority vested in the controller) and, significantly, includes a right to object to profiling based on those provisions. If a data subject objects to the processing of their data, the controller must 'no longer process the data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims'.[145] A data subject who objects to processing on this ground also has a right to restrict processing.[146] Significantly, he or she also has the right to require 'the erasure of personal data concerning him or her without undue delay' if 'there are no overriding legitimate grounds for the processing'.[147]

Article 21(2) also provides data subjects with a right to object at any time to the processing of their data for direct marketing purposes 'which includes profiling to the extent that it is related to such direct marketing'.[148] A data subject who objects to processing on this ground also has 'the right to obtain from the controller the erasure of personal data concerning him or her without undue delay'.[149]

Third, and most significantly, art 22 contains a right not to be subject to automated decision making, 'including profiling, which produces legal effects concerning … or similarly significantly affects' the data subject.[150] In order to come within the scope of art 22(1), the profiling in question must produce 'legal effects' (ie affect a person's legal status) or 'significantly affect' the data subject. This means that it does not apply to what Ustaran has described as 'common profiling' (ie profiling that simply 'involves analysing or predicting aspects of someone's life').[151] The extent to which an effect qualifies as significant remains to be determined.[152] However, recital 71 provides examples of decisions significantly affecting a data subject including 'automatic refusal of an online credit application or e-recruiting practices without any human intervention'.[153]

Three exceptions are provided for in art 22(2). The right not to be subject to automated decision making does not apply if the decision:

[144] Ibid art 21(1).

[145] Ibid.

[146] Ibid art 18(1)(d).

[147] Ibid art 17(1)(c).

[148] Ibid art 21(2).

[149] Ibid art 17(1)(c).

[150] Ibid art 22(1).

[151] Eduardo Ustaran, *Profiling – Sense and Sensibility* (20 May 2017) LinkedIn <https://www.linkedin.com/pulse/profiling-sense-sensibility-eduardo-ustaran>.

[152] See also the Article 29 Working Party Guidelines which state that for data processing to significantly affect someone the decision must have the potential to: 'significantly affect the circumstances, behaviour or choices of the individuals concerned; have a prolonged or permanent impact on the data subject; or at its most extreme, lead to the exclusion or discrimination of individuals': Article 29 Data Protection Working Party, *Guidelines on Automated Decision-Making and Profiling*, above n 141, 21.

[153] *GDPR* [2016] OJ L 119/1, recital 71.

(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller [the 'contractual' exception];

(b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests [the 'authorised by law' exception]; or

(c) is based on the data subject's explicit consent [the 'consent' exception].[154]

The application of the exceptions is subject to two sets of safeguards: in the case of the contractual and consent exceptions, art 22(3) requires 'the data controller [to] … implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision'.[155] The safeguarding of the data subject's rights and freedoms and legitimate interests is also built into the authorised by law exception.[156]

Article 22(4) provides a further safeguard: it provides that automated decision-making permitted under any of the three exceptions may 'not be based on special categories of personal data'[157] unless the 'explicit consent'[158] of the data subject is obtained or it 'is necessary for reasons of substantial public interest',[159] and provided that in both cases 'suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place'.[160]

Also relevant is recital 71, which applies in respect of all permissible profiling. This requires controllers to

> use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect.[161]

G *Right to Erasure*

Another significant provision is art 17, which confers on data subjects the right to seek the erasure of their personal data inter alia in the following circumstances:

[154] Ibid art 22(2).

[155] Ibid art 22(3).

[156] Ibid art 22(2)(b).

[157] Ibid art 22(4).

[158] Ibid art 9(2)(a).

[159] Ibid art 9(2)(g).

[160] Ibid art 22(4).

[161] Ibid recital 71.

(i) Where retention of the data is 'no longer necessary in relation to the purposes for which they were collected or otherwise processed';[162] this obligation gives more teeth to the principle of data minimisation.

(ii) Where the only justification for collection of processing the data is based on consent and that consent has been withdrawn;[163]

(iii) Where the data subject has exercised their right to object to its processing and 'there are no overriding legitimate grounds for the processing, or the data subject objects to the processing' of data for direct marketing purposes.[164]

(iv) Where the data has been processed unlawfully;[165] this would include where there have been breaches of the notification requirement[166] or the requirement to ensure that any data processed is accurate.[167]

(v) Where the personal data is processed in relation to the offer of information society services to a child.[168]

VII THE PRODUCTIVITY COMMISSION'S RECOMMENDATIONS

In 2016 the Australian Productivity Commission received a reference to 'undertake an inquiry into the benefits and costs of options for increasing availability of and improving the use of public and private sector data by individuals and organisations'.[169] This reference, which was given in the context of broader government policy to improve the availability and use of public sector data,[170] included requirements to '[i]dentify options to improve individuals' access to public and private sector data about themselves' and to '[e]xamine ways to enhance and maintain individuals' and businesses' confidence and trust in the way data are used'.[171]

The Commission's Inquiry Report, which was published in March 2017, recommended the introduction of a new regime based on

[162] Ibid art 17(1)(a). Article 5(1)(c) requires that personal data must be 'adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation')'. See also *Privacy Act* sch 1 cl 11.2.

[163] *GDPR* [2016] OJ L 119/1, art 17(1)(b).

[164] Ibid art 17(1)(c). The rights of data subjects under art 21 are discussed above: see above Part VI(F).

[165] Ibid art 17(1)(d).

[166] See ibid art 13. See also *Privacy Act* sch 1 cl 5.

[167] See *GDPR* [2016] OJ L 119/1, art 5(1)(d). See also *Privacy Act* sch 1 cl 10.

[168] *GDPR* [2016] OJ L 119/1, art 17(1)(f).

[169] *Data Availability and Use: Inquiry Report*, above n 7, v.

[170] The terms of reference refer to the Public Data Policy Statement which formed part of the Australian Government's National Innovation and Science Agenda: see ibid vii.

[171] Ibid vi–vii.

1. a new right that enables both opportunities for active data use by consumers and fundamental reform in Australia's competition policy
2. a structure for data sharing and release that would allow access arrangements to be dialled up or down according to the different risks associated with different types of data, uses and use environments.[172]

The new comprehensive right would apply to consumer data which is digital data,

> provided in machine-readable format, that is:
>
> - held by a product or service provider, and
> - identified with a consumer, and
> - associated with a product or service provided to that consumer[173]

and

> would enable consumers to:
>
> - share in perpetuity joint access to and use of their consumer data with the data holder
> - receive a copy of their consumer data
> - request edits or corrections to it for reasons of accuracy
> - be informed of the trade or other disclosure of consumer data to third parties
> - direct data holders to transfer data in machine-readable form, either to the individual or to a nominated third party.[174]

Significantly, the recommendations in the Final Report exclude the rights to opt out of data collection and appeal automated decisions made by predictive algorithms, which had been suggested in the Draft Report.[175]

A key issue with the new comprehensive right is the extent of its overlap with rights available in respect of 'personal information' under the *Privacy Act* and whether it is logical to set up a parallel regime administered by a different regulator. The new right is broader in its coverage than the *Privacy Act* (primarily due to the operation of the 'small business operator' exemption in the latter)[176] and the rights attached to it are also somewhat broader. For example, the *Privacy Act* does not provide a specific right to be informed of trade or other disclosures and does not require the transfer of personal data in machine-readable format. However, it is unclear why it would not be preferable to amend the *Privacy Act* to address these issues and it is doubtful whether these changes, although valuable, would do much to address the identified problems posed by Big Data.

[172] Ibid 14.

[173] Ibid 36.

[174] Ibid 35.

[175] Ibid 18, 216; Productivity Commission, 'Data Availability and Use' (Draft Report, October 2016) 350 ('*Data Availability and Use: Draft Report*').

[176] *Privacy Act* s 6D.

The Commission's proposed new release scheme is to be implemented via a new Data Sharing and Release Act, which includes oversight by a new National Data Custodian, release of data by Accredited Release Authorities (mainly existing government bodies that already release data) and the accreditation of trusted users to access data under its control or governance.[177] This proposed regime is based on a two-tiered system; sensitive, identifiable data would only be shared with trusted users, while non-sensitive data would be readily accessible to the public via National Interest Datasets.[178] Controversially, the Commission also recommends abolishing current requirements to destroy linked datasets and statistical linkage keys when data integration projects have been completed,[179] and extension of the existing exceptions in the *Privacy Act* which facilitate the processing of personal data for health and medical research purposes to cover public interest research in general.[180] An aspect of this proposed regime which is open to criticism is that it assumes de-identification to be sufficient in terms of securing the anonymity of data, and that further legal protection is not required against the risk of re-identification.[181]

A regime involving trusted users provides a useful way forward in terms of allowing for the release of sensitive information for public interest but arguably requires better safeguards to be implemented. Moreover, it addresses only the issue of the release of government datasets.

VIII NEW OFFENCES FOR RE-IDENTIFICATION OF DE-IDENTIFIED DATA

The Privacy Amendment (Re-identification Offence) Bill 2016 (Cth), which was introduced into Parliament in the aftermath of decryption by researchers of aspects of 'de-identified' health information which had been made publicly available by the government,[182] seeks to amend the *Privacy Act* to include two new criminal offence provisions. These offences, which relate to re-identifying de-identified government data, and publishing or communicating any re-identified dataset, carry penalties of up to two years imprisonment and 120 penalty units ($25 200), or a civil penalty of up to 600 penalty units ($126 000).[183] These offences, which

[177] *Data Availability and Use: Inquiry Report*, above n 7, 40–2, 46–7.

[178] Ibid 315.

[179] Ibid 44.

[180] Ibid 43.

[181] See, eg, ibid 42.

[182] Chris Culnane, Benjamin Rubinstein and Vanessa Teague, *Understanding the Maths is Crucial for Protecting Privacy* (29 September 2016) Pursuit <https://pursuit.unimelb.edu.au/articles/understanding-the-maths-is-crucial-for-protecting-privacy>. This research was conducted with a view to testing the security of the anonymisation techniques used to de-identify the data and its results were notified to the government.

[183] Privacy Amendment (Re-identification Offence) Bill 2016 (Cth) sch 1 cl 5. The same conduct, as well as a failure to notify, is also the subject of civil penalty provisions.

are intended to operate retrospectively from 26 September 2016, apply only to government datasets that are made generally available to the public.[184]

The proposed Bill, which seems to have stalled for now, has been criticised on the basis that it potentially criminalises white-hat hacking, the decryption of data with a view to exposing weaknesses in its anonymisation, and thereby discourages 'investigation and research into information security'.[185] From the perspective of Big Personal Data, it is open to criticism on the basis of its limited coverage. The proposed offences would apply only to government datasets that are made generally available to the public and are confined in their application to Australian government agencies and the private sector (including small business operators). As pointed out by the Australian Information Commissioner,

> the majority of acts, practices, and/or organisations which are currently exempt from the application of the *Privacy Act* will also be exempt from the scope of the Bill. Acts or practices currently exempt from the *Privacy Act* include acts done by media organisations in the course of journalism; political acts and practices; and, as most Commonwealth legislation (including the *Privacy Act*) does not bind the States and Territories, the activities of state and territory bodies (including their employees) are also exempt. I note that the majority of universities in Australia are State and Territory bodies.[186]

It would seem that the Bill is intended to deal with the situation where government releases datasets that are inadequately anonymised. Arguably a better way forward would be to adopt the Information Commissioner's recommendation that agencies need to implement practices, procedures, and systems to ensure that they comply with the *Privacy Act*, including 'taking reasonable steps to ensure personal information is not disclosed through open publication'.[187]

IX DATA BREACH NOTIFICATION

Also relevant are new requirements in pt IIIC of the *Privacy Act*, which commenced in February 2018.[188] Requirements to notify data subjects of breaches are now increasingly common overseas, including in the *GDPR*.[189] The notification requirement applies in circumstances where 'there is unauthorised access to, or unauthorised disclosure of, the information' (or such access or unauthorised disclosure is likely to occur) and 'a reasonable person would conclude that

[184] Ibid. This was the day on which the Attorney-General made the announcement of his intention to legislate a re-identification offence: see Explanatory Memorandum, Privacy Amendment (Re-identification Offence) Bill 2016 (Cth) 3 [10].

[185] Senate Legal and Constitutional Affairs Legislation Committee, Parliament of Australia, *Privacy Amendment (Re-identification Offence) Bill 2016* (2017) 8 [2.15].

[186] Timothy Pilgrim, Submission No 5 to Senate Legal and Constitutional Affairs Legislation Committee, *Privacy Amendment (Re-identification Offence) Bill 2016*, 3.

[187] Ibid 2.

[188] *Privacy Act* pt IIIC, as inserted by *Privacy Amendment (Notifiable Data Breaches) Act 2017* (Cth) sch 1 cl 3.

[189] *GDPR* [2016] OJ L 119/1, art 34.

the access or disclosure would be likely to result in serious harm to any of the individuals to whom the information relates'.[190]

The rationale for data breach notification requirements has been described as recognising 'that "individuals need to know when their personal information has been put at risk in order to mitigate potential identity fraud damages"'.[191] However, there is no evidence that they make a substantial difference,[192] and there are many instances when notification is of no assistance to data subjects in avoiding the potential harms arising from Big Personal Data.[193]

X OTHER POSSIBLE LEGAL APPROACHES

An alternative or supplementary way of handling some of the issues raised by end uses of Big Personal Data is to enhance the legal regimes that are designed to address similar issues. This may be necessary in light of the difficulties in amending data protection laws to fully address the issues raised by group privacy and re-identification.

Based on the analysis above, there are two key end uses that require attention — its use to discriminate against individuals in ways which are illegitimate or unjustifiable; and its use to manipulate or exploit behavioural insights in ways which are inappropriate or unfair.

The issue of discrimination has been highlighted by Citron and Pasquale in the context of what they describe as the 'scored society', a society in which decisions are reached on the basis of secretly developed scoring schemes; for example, those used in the context of decisions relating to credit applications.[194] They suggest that scoring results in discrimination on grounds which are prohibited under anti-discrimination laws[195] and also other forms of discrimination — discrimination that results from scoring which creates negative spirals that 'reinforce patterns of entrenched privilege and disadvantage'[196] and the discrimination that results from the arbitrariness of algorithms.[197] It is arguable that part of the solution to this problem may lie in expanding existing anti-discrimination laws to broaden

[190] *Privacy Act* s 26WE(2) (definition of 'eligible data breach').

[191] Australian Law Reform Commission, above n 62, vol 2 [51.7], quoting Canadian Internet Policy and Public Interest Clinic, 'Approaches to Security Breach Notification: A White Paper' (White Paper, Canadian Internet Policy and Public Interest Clinic, 9 January 2007) 2.

[192] There is limited data available concerning their efficacy but a study of US laws over the period 2002–9 suggested that they reduced the incidence of identity theft by 6.1 per cent: see Sasha Romanosky, Rahul Telang and Alessandro Acquisti, 'Do Data Breach Disclosure Laws Reduce Identity Theft?' (2011) 30 *Journal of Policy Analysis and Management* 256.

[193] See further David C Vladeck, 'Consumer Protection in an Era of Big Data Analytics' (2016) 42 *Ohio Northern University Law Review* 493, 504.

[194] Danielle Keats Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89 *Washington Law Review* 1.

[195] Ibid 14–15.

[196] Ibid 32.

[197] Ibid 25–6.

the range of protected attributes (for example, to address the group-based harms discussed above) and also the contexts in which discrimination is illegal.

In the case of behavioural manipulation, consumer protection laws provide the most obvious vehicle for reform, at least in relation to practices that affect individuals in their capacity as consumers. More generally, it is arguable that consumer protection law may have a broader role to play in addressing the autonomy-related harms to individuals arising from Big Personal Data.[198] As explained by Rhoen, 'Big [D]ata is shifting power away from consumers and data subjects towards data controllers' and enables them 'to limit the capacity of [data subjects] to act in their own best interest'.[199] He points out that consumer protection law is well suited to addressing this issue as it 'specifically aims to address power differentials based on information asymmetries in the market'.[200] While Australian consumer law has not to date played any significant role in relation to privacy protection, it is arguable that its prohibitions in respect of unfair contract terms and unconscionable conduct may have a useful role to play if further developed to address the consumer issues resulting from Big Personal Data. Other aspects of manipulation may call for sui generis solutions (for example, a reform of political advertising laws in the case of uses of Big Personal Data to manipulate election outcomes).

XI CONCLUSION

The *Privacy Act* falls short of providing adequate protection in respect of the privacy-related challenges posed by Big Data. There are a number of features of the *GDPR*, which if adopted in Australia, would arguably go some way towards addressing these issues. However, they do not address the significant issue of group privacy or fully address the issue of re-identification of data. Nor do they address the difficulties arising in a Big Personal Data context of applying the Australian Privacy Principles or relying on the consent model. Some European commentators have suggested that the shortcomings of the *GDPR* in addressing the challenges of Big Personal Data are not accidental. Mayer-Schönberger and Padova suggest that the *GDPR* incorporates 'important changes which directly address some of the key demands of the Big Data community',[201] while Gonçalves argues that 'caught between its twofold objective of strengthening the rights of the data subjects, and facilitating business, the EU legislator ended up favouring the latter to the detriment of the former'.[202] Meeting the challenges to privacy of

[198] In the case of the US, which lacks across-the-board privacy laws, the Federal Trade Commission has played a key role in regulating private sector privacy. For a comprehensive discussion of this role: see, eg, Andrew Serwin, 'The Federal Trade Commission and Privacy: Defining Enforcement and Encouraging the Adoption of Best Practices' (2011) 48 *San Diego Law Review* 809.

[199] Michiel Rhoen, 'Beyond Consent: Improving Data Protection through Consumer Protection Law' (2016) 5(1) *Internet Policy Review* 3, 3 <https://policyreview.info/node/404/pdf>.

[200] Ibid 6.

[201] Mayer-Schönberger and Padova, above n 124, 334.

[202] Gonçalves, above n 138, 114.

personal information posed by Big Personal Data is a highly complex matter and, in order for it to have any chance of success, the political will to address these challenges must be present.

In the Australian context, it is doubtful that implementation of the Productivity Commission's recommendations would substantially increase protection although they may provide a way forward in terms of the release of government datasets. Likewise, the new proposed offences for re-identification of anonymised data are very limited in their operation. It is arguable that as well as introducing an enhanced *Privacy Act* which includes protections extending beyond those provided for in the *GDPR*, Australian privacy measures might most usefully be supplemented via laws which have the potential to address the two key issues arising from end uses of Big Personal Data — discrimination and behavioural manipulation. Again, the extent to which Australian policy makers are prepared to intervene to regulate the deployment of Big Data by both the public and private sectors will be crucial in determining the success or otherwise of measures aimed at protecting Australians against the challenges posed by Big Personal Data.