DAVID SMITH

THE CITIZEN AND THE AUTOMATED STATE: EXPLORING THE IMPLICATIONS OF ALGORITHMIC DECISION-MAKING IN THE NEW ZEALAND PUBLIC SECTOR

LAWS 591: MASTERS THESIS

FACULTY OF LAW

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

2019

Abstract

Algorithms increasingly influence how the state treats its citizens. This thesis examines how the New Zealand public sector's use of algorithms in decision-making brings benefits, but also invites risks of discrimination, bias, intrusion into privacy and unfair decision-making.

This thesis's central conclusion is that these risks require a new response. New Zealand currently has a patchwork of existing protections which provide some deterrent against poor algorithmic decision-making. The Privacy Act 1993, Official Information Act 1982, New Zealand Bill of Rights Act 1990, Human Rights Act 1993 and applicable administrative law principles can provide remedies and correct agencies' poor behaviour in certain cases. But important gaps remain. This thesis examines these protections to show that they do not adequately stem cumulative and systemic harms, and suffer from important practical drawbacks. They do not provide the sound preventative framework that is needed; that is, one which ensures good public sector practice.

This thesis proposes a new regulatory model for public sector use of algorithms. It argues that a key element of any effective regulatory response is the use of "algorithmic impact assessments". These assessments would mitigate potential risks, and legitimise proportionate public sector use, of algorithms. It is also proposed that an independent regulator complements these assessments by issuing guidance, undertaking algorithm audits, and ensuring political accountability through annual reporting to Parliament. Agencies would have new obligations to disclose how and when algorithms are used in decision-making. Meanwhile, citizens would gain an enhanced right to reasons for algorithmic decisions affecting them and a right to human review. Together these measures would establish a model which would safeguard responsible and effective use of algorithms in New Zealand's public sector.

Word length

The text of this thesis (including abstract, table of contents, footnotes and bibliography) comprises approximately 49,906 words.

Subjects and topics

Algorithms, Judicial Review, Official Information, Human Rights Act, Privacy Act, Unreasonable Search and Seizure.

Acknowledgements

First, I would like to acknowledge my fantastic supervisor Nicole Moreham. Thank you for your invaluable guidance and support.

Thank you also to my friends and family. In particular, I wish to acknowledge Jon Duffy, Esther Watt and Simon Davies for acting as sounding boards as I have developed my thesis.

Thank you to Trade Me and Chapman Tripp for each showing patience and flexibility as my employers over the course of my study.

Last, but definitely not least, I want to thank my wife Jess Birdsall-Day for her unconditional love and support. Thank you for taking the load and for encouraging me when I have been in the thick of things.

Table of contents

I Chapter One: Introduction

A Overview

Humans are fallible and so too are the machines we create in our image. This thesis considers the increasing use of algorithmic tools to assist with the decisions the state makes about citizens. It looks at the benefits and harms that may arise. It also explores legal accountability frameworks for those who might be the subject of algorithmic decision-making, interrogating the remedies available when things go wrong and proposing a way forward to optimise how algorithms are used. More broadly, it asks: are algorithmic decisions fair? Are they any worse than normal human decision-making? Is there adequate recourse to challenge algorithmic decisions? And what could be done to balance the risks and benefits of using algorithms in the decisions of the state? These questions are explored in the context of New Zealand's public sector and its legal system.

The answers to these questions are complex and vary depending on the context and the methods by which algorithms are used. Algorithms are not inherently good or bad but a reflection of humans' choices about how we create and use them. When used well, algorithms may ameliorate the biases and mental lapses to which humans are naturally prone when making complex decisions – ultimately improving decisions' consistency and fairness.

However, the legal structures governing the use of algorithms need reform. The use of algorithms creates risks of bias and discrimination, of intrusions into privacy, and of a lack of transparency and natural justice in decision-making. In some areas where algorithms are used, these issues are only trivial; in others, rights as fundamental as an individual's physical liberty could be impacted. As this thesis reveals, a number of legal frameworks say something about these issues. However, the protections these frameworks provide are not always well adjusted for the contours of algorithmic harms and in some cases do not respond at all. This thesis outlines several changes to these frameworks that would improve this position.

However, even with these changes a broader response is required. Because of the limitations of rights-based remedies and the prospect of accumulative harm to citizens, this thesis proposes the framework for a new regulatory model. This model would legitimise public sector agencies' ("**PSA**") use of algorithms within clear guidelines and subject to ongoing review from an independent agency. Importantly, it would also facilitate harm

prevention, and legal and political accountability for decisions about how algorithms are used. Without a response along these or similar lines, only limited legal avenues exist to protect against harmful or inappropriate algorithmic decision-making.

B Structure

This thesis is divided into three parts.

1 Algorithms: how they work, risks and benefits, and the current state of regulation

Chapter two covers the first part of this thesis. This section surveys how algorithms work, their risks and benefits, and the level of scrutiny and regulation of algorithms in New Zealand and abroad.

This section describes how algorithms are, at their core, merely recipes or sets of instructions. However, increased computing power and new techniques now allow these recipes to be used for decision-making across a broad range of public services, from criminal justice to tax. The benefits of doing so can be immense. Algorithms can reveal valuable new insights and correlations, increase efficiencies, and facilitate more personalised services. Algorithms may also overcome humans' own flaws – helping to ameliorate bias, ensuring consistent treatment of comparable cases, and helping prevent humans' mental lapses. When combined with human judgment, algorithms create the opportunity for significantly improved overall decision-making.

However, chapter two also outlines why the increasing use of algorithms in decision-making is a cause for concern. Algorithms can amplify our own biases and create self-justifying, harmful and potentially discriminatory feedback loops. Moreover, algorithms can appear as objective, rather than a reflection of humans' subjective choices, leading humans to withhold their own judgment and abrogate their accountability over matters addressed by the algorithm. Algorithms – particularly machine learning algorithms – can also operate as opaque "black boxes",[1] creating natural justice issues for those subject to their decisions. Harms can arise when algorithms unfairly classify individuals into categories, including when these categorisations are used for other purposes downstream or are used in ways that undermine privacy protections.

---

[1] A term made prominent by Frank Pasquale. See Frank Pasquale *The Black Box Society: The Secret Algorithms that Control Money and Information* (Harvard University Press, Cambridge, 2015).

The last part of chapter two surveys the state of algorithmic regulation in New Zealand and abroad. It shows that while the potential for algorithmic harms is increasingly obvious, for the most part there has been little concrete action in New Zealand – despite a recent government report[2] indicating discrepancies in algorithmic best practice across PSAs, and some relevant agencies issuing useful voluntary guidance. Overseas, Canada has made the most significant progress in implementing a regulatory model for PSA automated decision-making.[3] The European Union's General Data Protection Regulation[4] ("**GDPR**") also provides some inspiration, while a number of reports and articles have also suggested ways to improve the use of algorithms. These overseas examples provide a starting point for the regulatory model proposed in the final part of this thesis.

2 Responsiveness of existing legal remedies

The second part of this thesis is made up of chapters three, four and five, and focuses on the individual remedies available to those who may have been affected by a PSA's decision about them involving an algorithm. The intention of this section is to provide a global view of the mechanisms available to deter and remedy the harmful use of algorithms, and to assess their adequacy. Throughout this part, real and hypothetical examples are used to highlight the scope and limits of these legal avenues as applied to algorithms.

First, chapter three considers the informational rights and remedies available to individuals subject to a PSA's algorithmic decision. It shows that the Privacy Act 1993 ("**Privacy Act**") potentially provides a useful remedy for an affected party – particularly where a PSA has failed to take reasonable steps to ensure it relies on accurate and complete information. However, the Privacy Act's effectiveness is limited by a number of definitional issues. This chapter recommends changes to ensure the Act's responsiveness is not stymied by these definitional issues. However more practical issues – such as small awards and poor access to justice via the Human Rights Review Tribunal ("**HRRT**") – remain a drawback.

This chapter also highlights the limitations of a right to reasons under the Official Information Act 1982 ("**OIA**") or the Local Government Official Information and

---

2 Statistics New Zealand and Department of Internal Affairs *Algorithm Assessment Report* (Wellington, 2018).

3 Canadian Government *Directive on Automated Decision-Making* (April 2019).

4 Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data [2016] OJ L119/1 [GDPR].

Meetings Act 1987 ("**LGOIMA**"). It proposes expanding this right, while creating a new right to know when one is subject to an algorithmic decision and a new right to human review.

Lastly, the chapter briefly considers the application of the right to be free from unreasonable search and seizure under s 21 of the New Zealand Bill of Rights Act ("**NZBORA**") to algorithm use cases. It suggests consideration be given to changing an existing presumption that the right is unlikely to apply to public places – one that would undermine the right's application to prominent areas of concern, such as the use of automated facial recognition technology.

Chapter four then considers the extent to which anti-discrimination protections in the Human Rights Act 1993 ("**HRA**") could respond to harms arising from PSA use of algorithms. It shows that a claimant will have the best chance of establishing prima facie discrimination if the decision-making was substantially automated (an algorithmically "**directed decision**"), but may struggle where any discriminatory algorithmic outputs are just one factor considered by the decision-maker (an algorithmically "**informed decision**"). Moreover, the HRA will not respond to low level harms caused by differential treatment. However, if a claimant can establish a prima facie case, a PSA may struggle to "justify" the discrimination unless the algorithm's workings are sufficiently transparent. However, like the Privacy Act, access to justice through the HRRT remains an issue.

Chapter five explores the responsiveness of judicial review to algorithmic decisions. It shows that a number of grounds of judicial review are potentially available due to common issues arising from the use of algorithms, such as inaccuracy of outputs, human automation bias and complacency, and the challenges of ensuring transparency with the use of algorithms. Grounds of review may include taking into account irrelevant considerations, failure to take account of relevant considerations, impermissible fettering of a decision-maker's discretion, material error of fact, procedural unfairness and a failure to provide adequate reasons. As with a claim under the HRA, the chance of success will typically improve if a decision is solely or substantially automated. However, judicial review does not require evidence of material harm to succeed – making it potentially less evidentially burdensome than the HRA. With this lower threshold to access remedies, judicial review claims have a better chance of addressing systemic harm caused by government practice, even if compensation will not be available.

3 A new regulatory model for public sector use of algorithms

Having interrogated the efficacy of key legal frameworks responding to algorithmic harms, the last part of this thesis argues New Zealand should implement a top-down model regulating how PSAs use algorithms. To justify this model, it draws parallel with environmental regulation to suggest that the patchwork of existing protections – even with the changes suggested above – will not protect adequately against cumulative harms from the use of algorithms.

The proposed model would achieve, among other things, proportionate oversight of PSA use of algorithms, protection from unacceptable harms, minimum standards of transparency and procedural fairness, and a process for political and legal accountability for how algorithms are used.

The proposed model rests on two pillars. First, an independent statutory regulator (the "**Algorithms Watchdog**") whose role would be to provide best practice guidance to PSAs and Ministers, to regularly audit PSAs to ensure appropriate processes are implemented to prevent algorithmic risks, and to report annually to Parliament to ensure public accountability and political transparency.

Second, PSA use of algorithms would be built around "algorithmic impact assessments" ("**AIA**"), which would consider among other things, privacy, human rights, and ethical risks associated with the use of algorithms for decision-making. Where an AIA suggested an algorithmic use poses a high risk of potential harm, its use case would require Ministerial sign off and would be subject to periodic review through a "sunset clause". This would again ensure political accountability for potentially difficult policy decisions. Once an AIA was completed or signed off by the relevant Minister, PSAs could operate the algorithm for decision-making within the bounds of the relevant use case. Additionally, PSAs would be obliged to outline their use of algorithms on their websites. The Algorithms Watchdog would be the custodian of a publicly accessible register including all decision-making algorithms used by the public sector, and their applicable AIAs.

This thesis suggests that, together, these changes would ensure that New Zealand has a fit for purpose and world-leading regulatory framework that it is proportional and optimally adjusted to encourage the safe use of algorithms by PSAs.

C Scope: what this thesis does and does not do

Before launching into chapter two, it is worth including some clarifying comments on the scope of this thesis.

First, the focus of this thesis is on the use of algorithms by government (widely construed) for the purpose of decision-making. It is does not address private-sector use of algorithms. The focus is intentionally limited to government because of individuals' unavoidable entanglement with the arms of the state, the state's unique ability to mobilise legitimate coercive power for investigation and enforcement purposes, and – from a practical perspective – in order to do the subject matter justice within the restricted confines of a Masters thesis. This is not to say that there are not real harms that can arise from the private sector's use of algorithms which need addressing; witness the rise of "digital redlining", the influence of "echo chambers" and "filter bubbles", and the pernicious effects of digital targeting of the most vulnerable.[5] These matters are simply beyond the scope of this thesis.

Policy-makers tasked with creating a cohesive response to harms arising from algorithms will need to look at both public and private actors. They will want to consider how a larger regulatory regime might address each in a consistent way. However, this thesis allows the reader to bring the microscope squarely onto the particular challenges of state use of algorithms. Moreover, the difference in relationship between citizens and the state, and consumers and private businesses, demands a bifurcated analysis. While a private actor can do harm, it is unlikely to put you in jail or deport you because of an algorithmic analysis. Citizens' relationship with the state also gives rise to unique remedies – such as judicial review and those arising in connection with NZBORA – focused on, among other things, respect for human rights, natural justice and proportionality.

Second, this thesis intentionally focuses on avenues that will generally be most relevant to a PSA's use of an algorithm in decision-making and most important for an affected party. However, other remedies may be available. The Waitangi Tribunal is an obvious, quasi-legal route that has been used in the past in relation to the Department of Corrections' use

---

5 For more information, see Cathy O'Neil *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Penguin Books, New York, 2016); Virginia Eubanks *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor* (St Martin's Press, New York, 2017); and The Workshop *Digital Threats to Democracy* (May 2019).

of actuarial tools that affect Māori.[6] Where there has been egregious use of an algorithm in decision-making, a litigant may also be able to establish a duty of care owed by the PSA giving rise to a claim in negligence.[7] And, the Office of the Ombudsman provides a limited route to inquire into administrative decision-making.[8] In other cases, a statutory regime may provide for particular rights or remedies for a person affected by a PSA's decision, or the specific context may create grounds for other remedies. However, these examples are outside the scope of this thesis – instead this thesis focuses on those remedies which will be most widely applicable and/or responsive.[9]

Lastly: a word on terminology. Throughout this thesis there is liberal reference to the "public sector" and "public sector agencies" (PSAs), and occasionally to "the state" or "the government". These references are intended to generally refer to the public sector as defined by the State Services Commission; that is, a broad range of entities including the core public service departments, various statutory Crown entities, district health boards, and local government entities.[10] Elsewhere this thesis refers to "algorithmic decisions" and "algorithmic harms". For clarity, these short-hand expressions refer to decisions influenced by the use of algorithms, and harms which may arise from the use of algorithms in such decisions. And lastly, references to "citizens" should be read to apply to any other persons likely to be affected by a PSA's algorithmic decision, including permanent residents.

---

6 Waitangi Tribunal *The Offender Assessment Policies Report* (Wai 1024, 2005).

7 See *Couch v Attorney-General* [2008] NZSC 45, [2008] 3 NZLR 725.

8 See Ombudsmen Act 1975.

9 It is worth noting that while the Ombudsman can inquire into public sector conduct, it cannot enforce remedies and its processes turn slowly. This thesis's discussion of judicial review explores the administative law principles that will, in any case, be relevant to those inquiries.

10 See State Services Commission "What is the 'Public Sector'" (April 2018) < http://www.ssc.govt.nz/resources/what-is-the-public-sector/>. Some entities should, however, probably be excluded from this thesis's scope (e.g., state-owned enterprises and tertiary education institutions).

II Chapter Two: The Rise of Algorithmic Decision-making

A Algorithms and you

Algorithms are now a part of daily life, whether we like it or not.[11] Corresponding with an exponential increase in computing power and the rise of the internet over the last few decades,[12] algorithms now decide the ads we see as we browse the internet,[13] suggest new music or movies to consume,[14] determine the political messages we should be exposed to,[15] and help us quickly find the information we need.[16] Algorithms are also routinely used to determine individuals' access to credit and insurance,[17] to assess individuals' suitability for employment[18] and even to choose which players should be selected by professional sports teams.[19] For those with access to a device, it is hard to go a day without interacting with an algorithm in some way.

---

11 Balkin argues we are moving towards a society "organized around social and economic decision-making by algorithms, robots, and AI agents, who not only make the decisions but also, in some cases, carry them out". See Jack M Balkin "2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data" (2017) 78 Ohio St LJ 1217 at 1219. See also Lee Rainie and Janna Anderson *Code Dependent: The Pros and Cons of the Algorithmic Age* (Pew Research Centre, 8 February 2017) at 5.

12 S C Olhede and P J Wolfe "The Growing Ubiquity of Algorithms in Society: Implications, Impacts and Innovations" (2018) 376: 20170364 Phil Trans R Soc A 1 at 4.

13 See Joseph Turow *The Daily You: How the New Advertising Industry is Defining Your Identity and Your Worth* (Yale University Press, New Haven, 2012).

14 Netflix and Spotify are well-known examples of content streaming services that use algorithms to customise recommended content based on users' previous choices and other factors.

15 See Philip N Howard, Samuel Wooley and Ryan Calo "Algorithms, Bots and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration" (2018) 15 Journal of Information Technology & Politics 81; Freedom House *Freedom on the Net: Manipulating Social Media to Undermine Democracy* (November 2017); Alex Hern "How Social Media Filter Bubbles and Algorithms Influence the Election" *The Guardian* (22 May 2017); and The Workshop, above n 5 *Digital Threats to Democracy* (May 2019).

16 For example, Google's online search engine available at www.google.com.

17 The White House *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (Washington, May 2016) at 11 - 13.

18 The White House *A Report on Algorithmic Systems* at 13 - 16; House of Commons Science and Technology Committee *Algorithms in Decision-making: Fourth Report of Session 2017-19* (15 May 2018) at 18 - 20.

19 See Michael Lewis *Moneyball: The Art of Winning an Unfair Game* (W W Norton & Company, New York, 2004).

Perhaps more profound, however, is the increasing use of algorithms by governments and what this means for the relationship between the state and its citizens. Governments across the developed world are deploying algorithmic tools for state-provided services in a quest to achieve efficiencies of cost and speed, more accurate decision-making, and better service provision.[20] Algorithms are used in areas like criminal justice, social welfare, tax and intelligence services, and New Zealand is no exception. If big data is the new oil, algorithms are the exploratory drills cramming the horizon.

What do these uses look like in practice, across jurisdictions? Algorithms are often used for criminal justice applications: automated facial recognition technology ("**AFR**") is now widely used by police in the United Sates for the purpose of crowd control and digital "line-ups" of suspected offenders.[21] Likewise, algorithms are used in sentencing and for risk assessments,[22] and to scan vehicle licence plates for traffic and criminal offences.[23] Algorithms are also used for predictive "hot spot" policing (which operates on the assumption that crime tends to follow geographical patterns),[24] and to analyse social media for "suspicious" behaviour[25] and to predict future offenders and victims.[26] Algorithms are also used in a range of other ways, for example to flag likely instances of reimbursement fraud for public health services,[27] to help process passports and visas,[28] to assist with

---

20 For example, Australian law explicitly authorises at least 11 different government departments to use computers for automated decision-making. See Simon Elvery "How Algorithms Make Important Government Decisions – and How That Affects You" *ABC News* (21 July 2017).

21 Georgetown Law Centre on Privacy & Technology *The Perpetual Line Up: Unregulated Police Face Recognition in America* (Washington, October 2016).

22 See *Ewert v Canada* 2018 SCC 30; Alexander Babuta, Marion Oswald and Christine Rinik *Machine Learning Algorithms and Police Decision-Making Legal, Ethical and Regulatory Challenges: Whitehall Report 3-18* (September 2018) at 6-7; Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks" *ProPublica* (23 May 2016); and Ellora Thadaney Israni "When an Algorithm Helps Send You to Prison" *The New York Times* (October 26, 2017).

23 Elizabeth E Joh "The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing" (2016) 10 Harv L & Pol'y Rev 15 at 22.

24 The White House *Big Data: Seizing Opportunities, Preserving Values* (Washington, May 2014) at 31; Babuta et al, above n 22, at 3.

25 See Joh, above n 23, at 25.

26 Andrew G Ferguson "Big Data and Predictive Reasonable Suspicion" (2015) 163 U Pa L Rev 327; Joh, above n 23, at 26; Babuta et al, above n 22, at 3.

27 The White House, above n 24, at 6.

28 See Statistics New Zealand and Department of Internal Affairs, above n 2, at 36 - 37.

allocation of health and transport services,[29] to provide tax refunds,[30] and to help employers to confirm the eligibility of newly hired employees to work in the country.[31]

To help us understand this general terrain, the chapter starts by looking at what algorithms are and how they work, their risks and benefits, and how algorithms are regulated in New Zealand and overseas.

B What are algorithms and how do they work?

In very simple terms, an algorithm can be described as "a series of logical instructions that show, from start to finish, how to accomplish a task".[32] For decades governments have used algorithms in this broad sense for administrative decision-making, but these have typically been simple models or "business rules" – for example, a checklist to ensure that if someone earns X, then Y is allocated to their tax bracket.[33] However, this thesis is particularly interested in algorithms which harness modern computing power to make or help make decisions about citizens.

To avoid any confusion, algorithms are not the same as "big data" – but they are part of the same phenomenon.[34] Most definitions of big data "reflect the growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data".[35] As such, big data often refers to the very large datasets (potentially from a wide range of sources) that can be processed by algorithms, as well as new information that is produced as outputs. When machine learning ("**ML**") algorithms are applied to large datasets and new correlations are discovered, this is called "data mining".[36]

Because it informs later discussion, it is important to step through the different kinds of

---

29 At 36 - 37.

30 At 37.

31 At 52.

32 Hannah Fry *Hello World: How to Be Human in the Age of the Machine* (Transworld Publishers, London, 2018) at 8. See also Osonde Osoba and Willam Wesler *An Intelligence in our Image* (Rand Corporation, Santa Monica, 2017) at 4.

33 See Statistics New Zealand and Department of Internal Affairs, above n 2, at 5.

34 As Balkin states, "when we talk about robots, AI agents, algorithms, we are also usually talking Big Data and Internet connection, just as when we talk about Big Data, we are also usually talking about the regulation of robots, AI agents, algorithms and AI agents that process it." Balkin, above n 11, at 1223.

35 The White House, above n 24, at 2.

36 Etham Alpaydin *Introduction to Machine Learning* (MIT Press, Cambridge, 2014) at 2.

algorithms and their qualities. Algorithms are typically used in at least four different ways depending on the task at hand:[37]

(1) *Prioritisation*: Prioritisation algorithms are designed to rank or order data so that certain desired information is presented first. For example, prioritisation is used with Google search results, and to help services like Netflix or Spotify rank how content is presented to you. Whatever is prioritised depends on the design of the algorithm.

(2) *Classification*: Classification algorithms tend to attach attributes to data so that it falls into different categories. For example, an algorithm may slice and dice data into different segments based on different factors (e.g., a person's age, location, or gender). Classification is often used by online advertisers who want to target particular groups of people.

(3) *Association*: Association algorithms make connections between different pieces of data on the basis that "these things go together". This technique is often also used in internet advertising to recommend products (e.g., you may be shown a product because "most people who like this book also bought this book").

(4) *Filtering*: Filtering algorithms are focused on isolating what is important. For example, they are used in speech recognition (by filtering out background noise) and in Facebook and Twitter feed results.

Some algorithms are human programmed, while others rely on machine learning (a form of artificial intelligence). Human programmed algorithms tend to follow sequential instructions or rules (i.e., if this, then this) programmed by a human.[38] Where executed through software, the internal logic of these algorithms can be made accessible in a language (code) that humans understand.

ML algorithms, on the other hand are software applications that "train" themselves to produce an outcome when fed with enough relevant data.[39] As Alpaydin describes: [40]

---

37 Fry, above n 32, at 8 - 10. Nicholas Diakopoulous "Accountability in Algorithmic Decision Making" (2016) 59 Communications of the ACM 56 at 57 - 58.

38 Fry, above n 32, 10 - 11.

39 Statistics New Zealand and Department of Internal Affairs, above n 2, at 8.

40 Alpaydin, above n 36, at 3.

> Machine learning is programming computers to optimise a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer programme to optimise the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data, or both.

Importantly for the purposes of this thesis, ML algorithms: (a) operate on the basis of *correlation* rather than causation; (b) assume that future outcomes will reflect the historical data they are trained on; and (c) can continually learn and change (to optimise for the desired outcome) without humans being able to understand their internal logic.[41] Some ML algorithms are "supervised" by humans to ensure they are optimising correctly, while others are not.[42] As we will see in the following chapter, these facts have important implications for the requirements of relevancy and transparency in administrative decision-making.

Finally, an important quality of many algorithms is that they indicate the statistical probability of a person having certain characteristics, rather than known actual characteristics. Across the world algorithms are often used by the state to either predict a person's future behaviour or categorise them as a particular type of person. Using statistical probability across the whole dataset, an algorithm might infer that you have committed tax fraud (for example, because 98 per cent of other people associated with the same data also have), whether or not this is actually true. This has important implications for fairness and natural justice, which are discussed further below.

C Benefits of algorithms

Now that we understand broadly what algorithms are and how they work, what are the

---

41 Alpaydin, above n 36, at 2 - 3; Babuta et al, above n 22, at 3. The understandability of machine learning algorithms may vary depending on type.

42 Edwards and Veale explain: "'Supervised Learning' takes a vector of variables, such as medical diagnosis, known as 'ground truth'. The aim of supervised learning is to accurately predict this ground truth from input variables in cases where we only have the latter. 'Unsupervised learning' is not 'supervised' by the ground truth. Instead, ML systems try to infer structure and groups based on other heuristics, such as proximity." See Lilian Edwards & Michael Veale "Slave to the Algorithm? Why a 'Right to an Explanation' is Probably not the Remedy you are Looking for" (2017) 16 Duke Law & Tech Rev 18 at 25.

benefits that make them attractive to governments? Depending on the use case, algorithms can reveal valuable insights and correlations, increase efficiencies, and overcome human bias. Algorithms also bring the prospect of more personalised services, do not suffer from humans' mental lapses, and when combined with human judgment, have the prospect of improving overall decision-making. Lastly, in some cases algorithms can actually be more auditable than human decision-makers.

1 Algorithms can reveal valuable insights and correlations

Algorithms allow large datasets to be crunched for anomalies or patterns that would not otherwise be obvious.[43] For example, the United States' Federal Trade Commission has reported how education institutions have used "big data techniques to help identify students who are at risk of dropping out and in need of early interventions strategies."[44] In the United Kingdom, the House of Commons' Science and Technology Committee has likewise noted how ML algorithms may help find patterns for diagnosing rare diseases when loaded with genomic data, and more generally can be used in the field of epidemiology to "detect and track infectious disease outbreak, […] enhance medical monitoring, and to optimise demand management and resource allocation in healthcare systems".[45] As an example, recent New Zealand research suggests that using a ML algorithm can substantially improve heart attack diagnosis.[46] These kinds of insights and benefits, often from data mining, would not necessarily be available without algorithms.

2 Algorithms can increase efficiency

Algorithms can increase efficiencies by carrying out roles that would otherwise have to be performed by humans. For example, algorithms can be deployed via chat bots to help provide answers to customers. They can also be used to attend to otherwise menial tasks or first-order decisions – such as data entry, assessment of various kinds of applications that government departments receive, or, in the education sphere, the grading of simple exam

---

43 The White House, above n 24, at 6 -7.

44 Federal Trade Commission *Big Data: A Tool for Inclusion or Exclusion – Understanding the Issues* (January 2016) at 6.

45 House of Commons, above n 18, at 11.

46 Michael Neilson "Artificial Intelligence Assists with Heart Attack Diagnosis in NZ-led Research" *New Zealand Herald* (11 September 2019).

scripts.[47] Algorithms also bring the potential for increased job satisfaction, by allowing workers extra time to focus on less tedious work.[48]

Decisions may also arrive more quickly than if they had to be fully processed by humans.[49] And most obviously, algorithmic decision-making can allow public sector agencies to save large sums of money that would otherwise be required to pay and support human staff, allowing the limited resources of the state to be spread more effectively.[50]

3 Algorithms bring the prospect of more personalised services

Algorithms may also contribute to more personalised services which better cater to individuals' needs. For example, the ability to analyse an individuals' specific data against a much larger dataset of well researched data could help doctors "more effectively perform 'precision medicine', an approach for disease treatment and prevention that considers individuals variability in genes, environment, and lifestyle".[51] Moreover, algorithms which can infer what most people in a certain group are likely to want, or how they generally behave, can suggest a pre-selected approach to suit a person's needs.

4 Algorithms can overcome bias

Humans are well known for exhibiting both actual and sub-conscious bias in their decision-making. Algorithms – when thoughtfully created and used – bring the prospect of overcoming these biases to improve human-decision making. Unlike humans, algorithms can operate emotionlessly on the "cold hard facts", without reference to factors which can influence humans (e.g., age, gender, race, religion).[52] For example, a risk assessment program used for bail in New Jersey led to a 16 per cent decrease in the pre-trial jail population – a figure that could have accounted for human biases that normally encourage

---

47 Emma Martinho-Truswell "How AI Could Help the Public Sector" *Harvard Business Review* (29 January 2018).

48 Ibid.

49 Martinho-Truswell, above n 47; Statistics New Zealand and Department of Internal Affairs, above n 2, at 4.

50 Statistics New Zealand and Department of Internal Affairs, above n above n 2, at 4.

51 Federal Trade Commission, above n 44, at 7; see also House of Commons, above n 18, at 11.

52 Joh, above n 23, at 28.

a conservative approach to bail.[53] In this case, the model was vetted to ensure that variables such as race, gender and other proxies (e.g. post code) were excluded. Likewise, big data tools could in theory help police to avoid using racial characteristics as proxies for suspicious behaviour.[54]

5 *Algorithms do not suffer from the mental shortcomings of humans and can improve decision-making*

Related to the previous point, algorithms may help to avoid a range of innocent mental phenomena and blindspots that plague human decision-making.

First, algorithms offer consistency. As far back as the 1950s and 1960s, it has been apparent that a professional's model for making decisions, when applied as an algorithm, tends to produce more consistently reliable results than the professional themselves.[55] Repeated research has shown that professionals will often contradict their own prior judgments when given the same data again.[56] As some scholars put it: "the problem is that humans are unreliable decision makers; their judgments are strongly influenced by irrelevant factors, such as their current mood, the time since their last meal, and the weather".[57] Algorithms, on the other hand, are consistent and do not tire.[58]

Second, algorithms bring the prospect of overcoming common mental traps that plague humans. Humans typically operate on two systems of thinking – a "system 1" which is humans' fast and instinctive way of responding to the world, and whose autopilot nature can explain the ability to arrive home without remembering the drive, and a "system 2" that involves a more concentrated, slower, and deliberative way of thinking.[59] While

---

53 Israni, above n 22.

54 Joh, above n 23, at 28.

55 See Lewis R Goldberg "Man versus Model of Man: A Rationale, Plus Some Evidence, for a Method of Improving on Clinical Inferences" (1970) 73 Psychological Bulletin 422; and Paul E Meehl *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (University of Minnesota Press, Minneapolis, 1954) at 119.

56 Daniel Kahneman, Andrew M Rosenfield, Linnea Gandhi and Tom Blaser "Noise: How to Overcome the High, Hidden Costs of Inconsistent Decision-Making" *Harvard Business Review* (October 2016) at 40.

57 At 40.

58 Babuta et al, above n 22, at 10; and Goldberg, above n 55, at 423.

59 See Daniel Kahneman *Thinking, Fast and Slow* (Penguin Books, London, 2012).

system 1 typically serves humans well, it can mean humans are more susceptible to a number of mental traps, including being "anchored" around a number or idea, [60] "availability bias" (e.g., thinking something is more likely because you recently saw an example), reasoning by "representativeness" (assuming something is "right" because it fits a stereotypical mental image of the thing), [61] and "hyperbolic discounting" (the natural tendency to disproportionately value things closer in time).[62] Some of these issues can in theory be overcome when decisions are made through the consistent logic of an algorithm.[63]

Last, algorithms can be tuned to catch patterns that humans struggle with. For example, while humans are often better at avoiding false positives in medicine, algorithms can be tuned to pick up everything that could meet the relevant criteria (avoiding false negatives). Studies of breast cancer screening have suggested using algorithms to put cancer screens in a "probably cancer bucket", and then having a human screen to reduce false positives, can bump up overall accuracy to 99.5 per cent.[64]

6 Algorithms can be more auditable than humans

Although the decision-making criteria of algorithms can be opaque, the same can be said for human decision-making. As Joh has described in relation to police decision-making, what stands out as suspicious in a person's mind can be the product of an "idiosyncratic, unaccountable, unknowable personal algorithm".[65] Even though a human can say why he or she made a decision, this may say little of the true drivers of his or her position.[66] In this

---

60 Amos Tversky and Daniel Kahneman "Judgement under Uncertainty: Heuristics and Biases" (1974) 185 Science 1124 at 1128.

61 At 1124 - 1127.

62 Allesandro Acquisti and Jens Grossklags "What can Behavioural Economics Teach Us about Privacy" in Allesandro Acquisti and others (eds) *Digital Privacy: Theory, Technologies, and Practices* (CRC Press, New York, 2007) 363 at 372; H Brian Holland "Privacy Paradox 2.0" (2010) 19 Widener LJ 893 at 905 - 906.

63 Kahneman et al, above n 56, at 44.

64 Fry, above n 32, at 90 citing Dayong Wang et al "Deep Learning for identifying metastatic breast cancer" Cornel University Library (18 June 2016). Unfortunately, Fry does not indicate the starting level of accuracy except to say pathologists correctly identify 96 per cent of straight-forward cases.

65 Joh, above n 23, at 29. For a counterview, see Eubanks, above n 5, at 168.

66 Oscar H Gandy, Jr "Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems," (2010) 12 Ethics and Information Technology 29 at 32.

sense, those the subject of human-programmed algorithmic decisions may be able to find a clearer explanation of why a decision was made – or at least a less self-serving one – than would be the case by relying on the human decision-maker's explanation.[67]

D Risks and harms of algorithms

While there are obvious benefits to gain from the state's use of algorithms, there are a variety of potential harms that need to be carefully considered. As the following part traverses, algorithms can create self-justifying, harmful and potentially discriminatory feedback loops, can be imbued with a patina of objectivity that discourages human accountability, and can operate as black boxes that raise natural justice issues. Further, algorithms often struggle to account for characteristics and groups outside the statistical norm, can create harm by defining individuals, and can overcome the purpose of privacy protections.

1 Algorithms can create self-justifying and harmful feedback loops

Algorithms can create self-justifying feedback loops which cause harm to individuals.[68] For various reasons, algorithms can produce results that justify similar future results, whether or not those results are normatively desirable or correct. This can arise due to the algorithm being trained on bad data, being poorly constructed, and/or being inadequately monitored to ensure it is performing as intended. The result is often the perpetuation and entrenchment of existing social biases, and/or unfair or inaccurate results.

Discriminatory or unfair outcomes can commonly arise if the algorithm is trained on biased data. For example, the House of Commons' Science and Technology Committee has noted that if historical "training data" are fed into a ML recruitment algorithm, it will continue making hiring decisions reflecting humans' past prejudices that males are better candidates.[69] Because males are *in fact* more commonly selected as "good candidates", this becomes the algorithm's "ground truth" that justifies, and potentially compounds, discrimination against women in hiring decisions. As Jack Balkin states, "humans program the algorithms with data, whose selection, organisation, and content contains the residue of

---

67 Joh, above n 23, at 29.

68 O'Neil, above n 5.

69 House of Commons, above n 18, at 18 - 20.

earlier discriminations and injustices".[70] If various algorithms employed in everyday life (not least government) rely on biased data, these kinds of effects can solidify human prejudices without the same level of visibility as the discriminatory decisions of humans.

Even when an algorithm does not use biased data, it still requires constant attention to ensure it does not embed unfair or discriminatory effects.[71] This is particularly the case with ML algorithms, which can drift as they "learn" with the more data they are fed.[72] However, rule-based algorithms, too, can cause unfair or discriminatory outcomes if they rely on compromised or inadequate data,[73] or the model and its assumptions are not constructed with care.

A well-known example is Equivant's COMPAS tool used in various states in the United States to provide predictive scores for offenders, including their likelihood of recidivism generally and their risk of violent recidivism.[74] Unlike our recruitment example, the tool does not include the discriminatory variable as an explicit criterion (in this case ethnicity), but nevertheless appears to penalise black offenders. A 2016 Propublica report analysed predicted and actual outcomes for over 10,000 offenders in Florida and found that while the tool had a similar level of accuracy for predicting re-offending as between black (63 per cent) and white offenders (59 per cent), its risk predictions created a high number of false positives that unfairly penalised black offenders compared with whites.[75]

The COMPAS example reveals the limitations of algorithms as predictive decision-makers. First, it shows that even when sensitive criteria such as race are excluded, proxies which correlate with these criteria (e.g., area code or employment status) can lead to decision-making which diverges based on such categories.

Second, the example shows that where an algorithm relies on existing data as an objective measure and there are disparities between groups, it cannot be both mathematically "fair"

---

70 Balkin, above n 11, at 1223.

71 At 1223.

72 Israni notes that if machine learning algorithms "are not constantly retrained, they 'lean in' to the assumed correctness of their initial determinations, drifting away from both reality and fairness". See Israni, above n 22.

73 See House of Commons, above n 18, at 20.

74 See Jeff Larson, Surya Mattu, Lauren Kirchnet and Julia Angwin "How We Analysed the COPMAS Recidivism Algorithm" *Propublica* (23 May 2016).

75 Angwin et al, above n 22; and Larson et al, above n 74.

in the sense of showing exactitude and "fair" in the sense of permitting equitable exceptions.[76] Corbett-Davies et al explain the quandary:[77]

> If the recidivism rate for white and black defendants is the same within each risk category, and if black defendants have a higher overall recidivism rate, then a greater share of black defendants will be classified as high risk. And if a greater share of black defendants are classified as high risk, then… a greater share of black defendants who do not reoffend will also be classified as high risk.

The implication: unless algorithms are programmed to embed a normatively desirable bent towards other concepts of "fairness" at the expense of accuracy, algorithms will serve to entrench or exacerbate existing disparities.[78]

In principle, there is no reason algorithmic bias could not cause feedback loops that entrench discriminatory treatment in the New Zealand context. Suppose police used a ML algorithm to predict which areas in a city or town are most likely to be risky areas for future crime – in a way similar to the well-known place based policing tool Predpol (the "**place-based policing scenario**").[79] This tool follows the logic that crime happens in patterns, allowing police to patrol neighbourhoods which are most susceptible to crime. Now, imagine a world where this tool is used in a neighbourhood predominantly made up of one ethnic group – for example, Pacific Islanders. Because the neighbourhood is policed more than others (and consequently more crime is discovered) the neighbourhood appears riskier than others. An increased Police presence leads to over-enforcement of minor crimes (such as drug possession and shoplifting), which in turn increases the rate of imprisonment among Pacific Islanders. This could create a feedback loop encouraging further Police focus on this neighbourhood, and thereby differential treatment of Pacific Islanders, compared with the rest of the population.[80]

---

76 See Jon Kleinberg, Sendhil Mullainathan and Manish Raghavan "Inherent Trade-Offs in the Fair Determination of Risk Scores" (17 November 2016); and Solon Barocas and Andrew D Selbst "Big Data's Disparate Impact" (2016) 104 Calif L Rev 671 at 721.

77 Sam Corbett-Davies, Emma Pierson, Avi Feller and Sharad Goel "A Computer Program Used for Bail and Sentencing Decisions was Labelled Biased against Blacks. It's Actually Not That Clear." *The Washington Post* (17 October 2016).

78 Gandy, above n 66, at 34 - 35.

79 For more on place-based policy and Predpol, see Andrew D Selbst "Disparate Impact in Big Data Policing" (2017) 52 Ga L Rev 109 at 130 - 136.

80 Osoba and Wesler highlight how this can be represented mathematically, to show "the effect of increasingly criminalizing specific subpopulations and (more critically) generating more 'objective'

2 *Algorithms can appear objective and diminish human responsibility*

When algorithms are used for decision-making, the outputs can carry a patina of objectivity which may reduce the likelihood the results are challenged.[81] Why?

For a range of reasons, decision-makers (or those who judge appeals from an algorithmic decision) may be hesitant to encroach on the algorithm's decision. Balkin argues a "substitution effect" lies at the heart of this phenomena.[82] He suggests humans can believe the algorithm has superior qualities to the erstwhile decision-maker, potentially because of the apparently "scientific" nature of the algorithm, and that this can lead humans to overlook the algorithm's limited nature and capabilities.[83] Human decision-makers may also project agency onto the algorithm, allowing them to distance themselves from responsibility for the results.[84] One can assume this risk is likely to be greater where responsibility for the creation, implementation and operationalisation of the algorithm is shared across a range of human actors. Lastly, Balkin observes how the algorithm can deflect attention from the social relations creating the algorithm's results. An algorithm may produce results imposing a judgement on a person (e.g., that she is likely to default on a loan, or is a higher risk of offending) in a way that focuses humans on a binary question of the "good" or "bad" result, rather than the social factors which contribute to this "score".[85] In circumstances where mistakes and errors are inevitable, these factors can reduce the use of what might be called "good discretion" to overcome an algorithmic

---

data to support future biased enforcement decisions." See Osoba and Wesler, above n 32, at 14 - 15. See also Selbst, above n 79 at 130 - 136.

81 The White House, above n 24, at 46; Colin Gavaghan, Alistair Knott, James Maclaurin, John Zerilli and Joy Liddicoat *Government Use of Artificial Intelligence in New Zealand: Final Report on Phase 1 of the New Zealand Law Foundation's Artificial Intelligence and Law in New Zealand Project* (Wellington, 2019) at 35 and 39.

82 Balkin, above n 11, at 1224 - 1225.

83 For an instructive case, see Jay Stanley "Pitfalls of Artificial Intelligence Decisionmaking Highlighted in Idaho ACLU Case" (2 June 2017) American Civil Liberties Union <https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case>; and *K.W. v Armstrong* No. 14-35296 (9th Cir. 2015).

84 Joh notes that by "…applying big data analytics to digitized information, big data tools appear to provide an objective analysis of information. But discretionary decisions can play an important role in big data in ways that may not be obvious", for example, the model, the data, how it is displayed and where the tool is applied, while the data itself is often a product of discretionary decision-making. See Joh, above n 23, at 30 - 31.

85 For anecdotal examples, see Eubanks, above n 5, at 138 - 173.

decision which seems unfair or to provide a second chance.[86]

These general points mirror studies in the field of robotic automation and decision-making. These studies suggest that, as the degree of automation increases, so too does humans' complacency to the possibility the machine is wrong – particularly when automation is generally highly reliable.[87] As Onnasch et al note, the research suggests a trade-off in which:[88]

> … more automation yields better human-system performance when all is well but induces increased dependence, which may produce more problematic performance when things fail…

In particular, humans are less likely to respond to performance failures when the degree of automation "moves across the critical boundary from information acquisition and information analysis to action selection, the latter alleviating the human from some or all aspects of choosing an action".[89] For example, pilots using a highly automated flight planning system are less likely to spend time generating and evaluating alternative options, and are more likely to take a recommended flight path which is sub-optimal, than when pilots use a less automated system.[90] Likewise, in the context of an engine fire, 75 per cent of pilots are likely to accept an automated but incorrect recommendation to shut down an engine, compared to 25 per cent of pilots using a traditional paper checklist of potential errors.[91] Similar failures to identify automation failure in emergency situations have been exhibited by ship officers.[92]

However, it is worth noting a tension lies between making it easier to challenge an algorithm's outputs and ensuring the benefits of algorithms are realised. As alluded to

---

86 Joh, above n 23, at 32.

87 Raja Parasuraman and Dietrich H Manzey "Complacency and Bias in Human Use of Automation: An Attentional Integration" (2010) 52 Hum Factors: J Human Fact Ergon Soc 381 at 390; and Linda Onnasch, Christopher D Wickens, Huiyang Li and Dietrich Manzey "Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis" (2014) 56 Hum Factors: J Human Fact Ergon Soc 476.

88 Onnasch et al, above n 87, at 477.

89 At 485.

90 Parasuraman and Manzey, above n 87, at 392.

91 At 392.

92 Kayvan Pazouki, Neil Forbes, Rosemary A Norman and Michael D Woodward "Investigation on the Impact of Human-automation Interaction in Maritime Operations" (2018) 153 Ocean Engineering 297.

above, humans are prone to mental shortcomings and biases. Algorithms can help address some of these. But, ironically, if intervention against the algorithmic decision is seen as easy and insignificant, decision-makers might run the risk of falling into confirmation bias: only allowing appeals from the algorithm where a different outcome accords with their own sensibilities or biases.[93] While further research is required to determine how to manage this tension, chapter three outlines why, for now, a right to human review nevertheless remains a sensible option.

3 Algorithms can operate as "black boxes"

Algorithms can cause harm to the extent their logic, training data and decision-making criteria are opaque. This is particularly troubling for algorithms used by the public sector because of the coercive power of the state. Harm can arise if citizens are unable to understand and challenge the bases on which decisions about them are made, and this leads to disentitlement to important public goods or services, or differential treatment and stigmatisation.

Algorithms can be opaque in at least three ways. First, algorithms can lack *transparency.* This can occur because individuals will not always know algorithms are making decisions about them, limiting an individual's ability to ask the right questions about why a decision was made.[94] In this broad sense, the use of algorithms is not always transparent or easily discoverable, limiting the means for accountability. As Joh suggests in relation to algorithmic policing, "powerful big data tools can operate secretly and without public awareness in ways that cases of street police brutality cannot."[95]

Moreover, algorithms can be non-transparent in both a technical and a legal sense. As outlined above, although programmed algorithms can be understood by professionals, humans cannot understand a ML algorithm's logic and how this produced the ultimate output (and this logic can change over time). From a legal perspective, both kinds of algorithms may also be protected as confidential trade secrets which are unable to be

---

93 See Babuta et al, above n 22, at 12. The authors suggest caution in deploying algorithms in the area of criminal justice until research establishes how decision-makers are actually influenced by them.

94 For example, see Ali Winston "Palantir has Secretly Been using New Orleans to Test its Predictive Policing Technology" *The Verge* (27 February 2018); and, in relation to facial recognition technology, see Georgetown Law Centre, above n 21.

95 Joh, above n 23, at 30.

publicly scrutinised for fairness.[96] So, even if an individual knows an algorithm is being used, both technical and legal measures can limit the ability to interrogate how any algorithmic output was created.[97]

Algorithms also often lack *explainability*.[98] This is related to the idea of transparency, but asks: is there a description of how the algorithm works which can be readily understood by someone? Even if someone is not a data scientist or software developer, and will not be able to understand the code, is there a sufficiently detailed and understandable explanation that would allow someone to argue that his or her treatment has been unfair?[99]

Lastly, algorithms should be *auditable* – but frequently are not.[100] The results of the algorithm should be able to be validated to see if the model is in fact working. Audits should be regular and frequent, or individuals may suffer avoidable harm. Auditability is also important from an accountability perspective – giving a potential plaintiff the chance to prove the unreliability or unfairness of the algorithm.[101]

Too much opacity – a lack of transparency, explainability, and auditability – tends to indicate poor algorithmic hygiene, which can create risks to individuals. As the following chapters discuss, this also leaves PSAs potentially vulnerable to legal challenge.

4 Algorithms can struggle to account for outlier groups and characteristics

When algorithms make predictions about, or classify, people, they rely on statistical

---

96 Both the Privacy Act and OIA provide possible exceptions to disclosure of information to protect trade secrets. See Privacy Act 1993, s 28, and Official Information Act 1982 [OIA], s 9(2)(b).

97 See Jenna Burrel "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms" (January – June 2016) Big Data & Society 1 at 3 - 5.

98 See Frank Pasquale "Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society" (2017) 78 Ohio St LJ 1243 at 1252.

99 Bruno Lepri, Nuria Oliver, Emmanuel Letouze, Alex Pentland and Patrick Vinck "Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges" (2017) 31 Philos Technol 611 at 621 - 622.

100 Christian Sandvig, Kevin Hamilton, Karrie Karahalios and Cedric Langbort "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms" (paper presented to the Data and Discrimination: Converting Critical Concerns into Productive Inquiry preconference of the 64th Annual Meeting of the International Communication Association, Seattle, 22 May 2014).

101 Edwards & Veale, above n 42, at 43. The authors note that that auditing can create accountability without necessarily needing complete technical transparency.

probability. However, this naturally means that groups which fall outside of the norm are particularly likely to suffer misclassification or mistreatment. Typically, these harms arise because a sample size, while satisfactory for the broader population, is too small for a particular group to provide valid results.[102] However, it can also arise because the product or service is aimed at a particular average customer and there is no intention to cater to outliers.

The most obvious examples relate to ethnic or cultural minorities that form a small proportion of the total population. For example, Google has mistakenly classified black people as gorillas,[103] driverless cars have struggled to detect dark-skinned pedestrians,[104] and gaming platforms have banned people who use "non-traditional" sounding names.[105] In a different context, a Canadian prisoner of Métis ethnicity sued the state on the basis that various psychological and actuarial risk assessment tools were developed for a predominantly non-indigenous population, and that no research confirmed that they were valid when applied to indigenous persons.[106] Although no evidence was brought to prove the tools were unsuitable, in *Ewert v Canada* the plaintiff succeeded because the Correctional Service of Canada failed to meet a statutory responsibility to take all reasonable steps to ensure information about the offender was accurate, complete and up to date.

Similar issues could theoretically arise in the New Zealand public service. In fact, a 2005 Waitangi Tribunal report examined a very similar question to *Ewert*, by asking whether the Department of Corrections' "Risk of Reconviction / Risk of Re-Imprisonment" ("**ROC*ROI**") tool – used in sentencing, probation and for parole hearings – might unfairly penalise Māori.[107] That tool does not include race as an explicit variable. It

---

102 ML algorithms can "have difficulty capturing specific cultural effects when the population is strongly segmented" and because ML algorithms are statistical estimation methods "their measures of estimation error often vary in inverse proportion with data sample sizes". See Osonde and Wesler, above n 32, at 19.

103 Israni, above n 22.

104 Sigal Samuel "A New Study Finds a Potential Risk with Self-Driving Cars: Failure to Detect Dark-skinned Pedestrians" *Vox* (6 March 2019).

105 Osonde and Wesler, above n 32, at 20. Algorithms have also been used to discriminate online based on "black sounding" names. See Latanya Sweeney "Discrimination in online ad delivery" (2013) 11 ACM Queue 10.

106 *Ewert v Canada*, above n 22.

107 Waitangi Tribunal, above n 6. See also Joel McManus "Why a Pastor who Abused Children Served Half as Much Prison Time as a Low-level Cannabis Dealer" *Stuff* (13 August 2019).

measures the probability that an individual offender will be reconvicted for new offending within 5 years[108] and relies on:[109]

> … the mathematical relationship between basic social and demographic variables (e.g. age, gender), criminal history variables (e.g. age of first offence, time free in community since thirteenth birthday, seriousness of previous offences, length of time between offences) and future offending.

At the time, the Tribunal did not consider on the evidence provided that it caused prejudice to the Māori claimants. [110] However, it could not rule out the prejudicial effects of another psychological tool (CNI/MaCRNs) because no reliable evidence existed about its effects, and called for urgent action in the use of actuarial tools to avoid possible prejudice. [111]

ROC*ROI has been applied to New Zealand conditions for over 20 years now and Corrections has a "high confidence in its accuracy".[112] But, one can imagine the risks of bias to particular groups in a slightly different scenario – say where a similar tool developed for the US market is bought "off the shelf" and used in New Zealand without appropriate adjustment to local demographic groups (the "**sentencing-bias scenario**").

5 Algorithms can unfairly classify people and cause cumulative harms

Related to the previous point, algorithms are frequently used to classify and define individuals. This has led Citron and Pasquale to speak of the "scored society".[113] While categorisation is not new, its scale (and the permanence of its markings) are, and the consequences can be significant. The harms that can arise from the classification of individuals are the flip-side to the benefits of personalisation that algorithms can create.

First, harms can arise when individuals are misclassified (e.g., because their characteristics are an outlier within the model). This is a particular danger for ML algorithms which are designed to detect correlation, but are unable to assess whether a correlation is associated

---

108 Statistics New Zealand and Department of Internal Affairs, above n 2, at 36.

109 Department of Corrections "Risk of Reconviction" < https://www.corrections.govt.nz/resources/research_and_statistics/risk-of-reconviction.html>

110 Waitangi Tribunal, above n 6, at viii.

111 At viii.

112 McManus, above n 107.

113 Danielle Keats Citron and Frank A Pasquale "The Scored Society: Due Process for Automated Predictions" (2014) 89 Wash L Rev 1.

with causation.[114] This can cause damage to reputation or the stigma that an individual is a "risky" kind of person.[115] The consequences can also be more immediate and practical: for example, a Boston man who was classified as holding a fraudulent drivers' licence was unable to get any job offers until the matter was addressed.[116] Other serious examples can include disentitlement to government provided support.[117]

Individuals can even have their assets seized due to algorithmic failures. As at the time of writing, thousands of current and former beneficiaries in Australia have had money automatically taken from their accounts as debts owing to the government, with the average amount being AUD 2,148.[118] These deductions have been routinely overturned in Australia's Administrative Appeal Tribunal due to inaccuracy, and the Australian Government has never appealed these decisions.[119] As a result, a class-action claim in unjust enrichment is now underway for these "robo-debts".[120]

However, even when a person is classified correctly, other harms can arise. First, a classification for one intended purpose can quickly become used for another. For example, in some United States' jurisdictions, the results of tools used to classify offenders' needs for rehabilitation purposes have also been used as proxies for offending risk. Napa County Superior Court Judge Mark Boessenecker suggests this approach can mean a person committing child sex offences every day:[121]

> … could still come out as a low risk because he probably has a job… meanwhile, a drunk guy will look high risk because he's homeless. These risks don't tell you whether the guy ought to go to prison or not; the risk factors tell you more about what the probation conditions ought to be.

Because of the range of algorithmic classifications which exist, decision-makers can also take another algorithm's outputs at face value and incorporate them into further algorithmic decisions without interrogating their integrity. As such, primary classifications (whether

---

114 Babuta et al, above n 22, at 21.

115 Balkin, above n 11, at 1238 - 1239.

116 Diakopoulous, above n 37, at 57.

117 Stanley, above n 83.

118 See "Robo-debt Class Action Could Deliver Justice for Tens of Thousands of Australians Instead of Mere Hundreds" *The Conversation* (17 September 2019).

119 Ibid.

120 Ibid.

121 See Angwin, above n 22.

accurate or not) can lead to secondary inferences by different decision-makers. This creates the danger that initial algorithmic classifications lead to negative path dependency where reliance on first-order results leads to cumulative algorithmic harm to an individual.[122]

Moreover, in a future world where individuals are regularly judged across a range of areas by the state, individuals may be inclined towards self-restraint in a way that prevents the full exercise of democratic rights. A mildly Orwellian real-life example is China's early steps toward a social credit system. Simplifying greatly, this system is designed to record and reward citizens who exhibit the Communist Party's view of socially desirable behaviour (e.g., giving to charity) and to penalise those who behave in less desirable ways (e.g., through traffic violations).[123] Meanwhile, Zimbabwe is using algorithms to enable "massive real-time video surveillance, facial recognition technology for the whole population, including other biometric identification programs".[124] On a very different scale, a recent Welsh case has also legitimised the use of AFR by police in public places,[125] which we could expect to affect people's behaviour in public. Even without widespread surveillance, if algorithms are used widely in Western countries to sort and make decisions about citizens, including their entitlement to certain state benefits, one can imagine how citizens may change their "identity, behaviour, or other aspects of personal self-presentation in order to appear less risky".[126] Moreover, as outlined below in chapter three, individuals may potentially perceive automated decision-making about them as demeaning and impacting on their dignity.[127]

Consider another hypothetical example of how the "scoring" and surveillance of citizens could occur in New Zealand. Imagine that the Ministry of Social Development ("**MSD**") has decided to use an algorithmic software for the purpose of determining whether someone is in a de facto relationship and hence, entitled to the Sole Parent Support benefit (the "**benefit surveillance scenario**"). The software uses information from MSD about those

---

122 See Gandy, above n 66, at 37 - 38.

123 Nicole Kobie "The Complicated Truth About China's Social Credit System" *Wired* (21 January 2019); and Zheping Huan "All Chinese Citizens Now Have a Score Based on How Well We Live and Mine Sucks" *Quartz* (10 October 2015).

124 "Mnangawa Invests in Super Spyware for Citizens" *Zambezi Post* (29 June 2019).

125 *R (Bridges) v CCSWP and SSHD* [2019] EWHC 2341 (Admin).

126 Balkin, above n 11, at 1238. See also Neil M Richards "The Dangers of Surveillance" (2013) 126 Harv L Rev 1934.

127 Min Kyung Lee "Understanding Perception of Algorithmic Decisions: Fairness, Trust and Emotion in Response to Algorithmic Management" (January – June 2018) Big Data & Society 1 at 12.

receiving the Sole Parent Support benefit (e.g., name, age, gender and location) and performs a search of information about the person publicly available on the internet. The software collects information from social media sites (such as Facebook, Instagram and Twitter) which may indicate social connections and whether the person is in a relationship. People are then sorted into buckets: red (for "likely") orange (for "maybe") and green (for "unlikely"). If MSD staff took the "red" results at face value – a risk outlined above – even an algorithm with a 95 per cent accuracy would lead to unfair downstream consequences for a large number of inaccurately classified individuals. Therefore, harm could arise first due to interventions based on algorithmic misclassification. Secondly, this surveillance could well change how beneficiaries express themselves online and/or disincentivise the beginnings of desirable relationships.[128]

6 Algorithms can undermine the objectives of privacy protections

Lastly, algorithms fed on big data can erode general privacy protections – particularly those that operate on the basis of association and/or classification. As Baracos and Nissenbaum note, ML algorithms fed on large enough datasets can produce generalisable rules that can infer private facts about individuals.[129] Persons who have not consented to the disclosure of this information can be "discovered" and that information used against them. Infamously, United States retailer Target was able to use data mining techniques to identify pregnant customers and send them baby-related advertising, sometimes in situations where others in the family did not know this fact (and were upset to find out).[130]

While PSAs often have compulsory information gathering powers for criminal and regulatory enforcement, those agencies are usually also subject to data protection laws (such as the Privacy Act) that limit the ability to collect and use citizens' information without consent, a legitimate purpose or in ways that could be considered unfair. However, by using algorithms to produce big data analytics, PSAs could uncover information individuals would refuse to provide, subverting individuals' rights to control how information about them is used.

---

128 While individuals no doubt already change their behaviour to respond to monitoring by PSAs, the scale and extent of monitoring possible with algorithmic tools creates new risks for individuals' expression.

129 Solon Barocas and Helen Nissenbaum "Big Data's End Run Around Procedural Privacy Protections" (2014) 57 Communications of the ACM 31 at 32.

130 At 32.

E Algorithmic regulation in New Zealand and abroad

Having explored the risks and benefits of using algorithms to make public-sector decisions, we now turn to the state of regulation in New Zealand and abroad.

1 The New Zealand landscape

This part explores how, while there is increasing awareness of potential and actual use of algorithms in by government in New Zealand, few real regulatory steps have been taken. It shows that, as a result, those affected by algorithmic decisions are limited to a range of traditional legal actions (discussed in the following chapters). While, thankfully few New Zealand examples indicate severe risks, and some agencies are now creating guidance for algorithmic tools,[131] there is no *preventative* framework other than PSAs' voluntary steps towards good practice.

First, New Zealand PSAs' use of algorithms is increasingly in the spotlight. While most concern to date has focused on the private sector's use of algorithms (think targeted advertisements and social media "filter bubbles"[132]), the New Zealand Government's use of algorithms has not been immune from bad press: in 2017 the Accident Compensation Corporation ("**ACC**") was reported to be using an algorithm that ran on client data without clients' consent or vetting from the Privacy Commissioner to "make predictions about which clients need more help, what type of case managers they should have, and how long they are likely to take to recover".[133] In 2018, reports emerged that Immigration New Zealand was using a model to determine which people were likely to "most commonly run up hospital costs or commit crime" for the purpose of deportation decisions, raising concerns that the model could discriminate based on age, gender or ethnicity.[134] Subsequently, it emerged that the model consisted of an excel spreadsheet that created a simple points system without evidence of statistical integrity.[135] And recently, the Privacy

---

131 See Statistics New Zealand and Department of Internal Affairs, above n 2.

132 The Workshop, above n 5, at 17.

133 Kirsty Johnston "Privacy and Profiling Fears over Secret ACC Software" *New Zealand Herald* (15 September, 2017).

134 Gill Bonnett "Immigration NZ Using Data System to Predict Likely Troublemakers" *Radio New Zealand* (5 April 2018); and Asha McLean "NZ Immigration Rejects 'Racial Profiling' Claims in Visa Data-modelling Project" *ZDNet* (6 April 2018).

135 Tze Ming Mok "Crap Models and Laughable Claims: Immigration NZ's Spreadsheet Fiasco" *The*

Commissioner raised concerns over the suggestion Auckland Transport might utilise widespread AFR.[136]

Outside of these examples, focus has typically fallen on the previous National Government's pursuit of the "Social Investment Approach". This approach looked to leverage data analytics to identify the most vulnerable members of society for intervention by government social services, and particularly children so as to address "problems at an earlier, more tractable stage".[137] However, it was not without controversy: MSD was chastised by the Privacy Commissioner for attempting to force social service providers to require their clients to provide personal information that could be shared with government for interventions related to this approach.[138] Likewise, a model created to identify children most likely to suffer harm or abuse[139] was – despite peer review[140] and accuracy comparable with mammogram screening[141] – criticised for creating a high chance of harmful false positives and for raising ethical issues for social service providers.[142] When MSD proposed to validate the accuracy of this predictive risk model for vulnerable children by choosing not to intervene in cases identified as high risk, the then Minister blocked the proposal on the basis the affected children were not "lab rats".[143] The same model has since been used controversially in Allegheny County in Pennsylvania.[144]

Spinoff (10 April 2018).

136 "Privacy Commissioner Would be 'Very Worried' if Auckland Transport Introduces Facial Recognition Technology in Cameras" *1News* (13 August 2019).

137 Office of the Prime Minister's Chief Science Advisor *Using Evidence to Inform Social Policy: the Role of Citizen-based Analytics* (Auckland, 19 June 2017) at 23.

138 Office of the Privacy Commissioner *Inquiry into the Ministry of Social Development's Collection of Individual Client-Level Data from NGOs* (4 April 2017).

139 See Rhema Vaithianathan, Tim Maloney, Nan Jiang, Irene De Haan, Claire Dale, Emily Putnam-Hornstein and Tim Dare *Vulnerable Children: Can Administrative Data be Used to Identify Children at Risk of Adverse Outcomes* (September, 2012). This model sprung from the government's work to reduce harm to "vulnerable children". For more information, see Ministry of Social Development *The White Paper for Vulnerable Children: Volume I* (October 2012) at 9 - 10; and Ministry of Social Development *The White Paper for Vulnerable Children: Volume II* (October 2012) at 75 - 81.

140 Tim Dare *Predictive Risk Modelling and Child Maltreatment: An Ethical Review* (25 September 2013).

141 Children identified by the model represented 37 per cent of all children in New Zealand who would receive maltreatment by age 5. See Vaithianathan et al, above n 139, at 3.

142 See Emily Keddell "The Ethics of Predictive Risk Modelling in the Aotearoa/New Zealand Child Welfare Context: Child Abuse Prevention or Neo-liberal Tool?" (2014) 35 Critical Social Policy 69.

143 Stacey Kirk "Children 'Not Lab-rats' – Anne Tolley Intervenes in Child Abuse Experiment" *Stuff* (30 July 2015).

144 Eubanks, above n 5, at 127 - 173.

The cross-government view of how algorithms are used has become clearer after an algorithmic "stock take" was undertaken by Statistics New Zealand and the Department of Internal Affairs in 2018 ("**Stocktake Report**").[145] The Stocktake Report reveals algorithms employed for a range of uses, varying from passport facial recognition technology, to ACC's fraud detection system, to tools intended to predict youth at greater risk of long term unemployment, to the ROC*ROI mentioned above. It focuses on government's use of "operational algorithms" (i.e., those used to help government make decisions) and suggests that the algorithms used across the public sector mostly pose low risk of harm to citizens and generally involve human final review and decision-making.[146] However use of algorithms is only likely to increase. Moreover, while the Stocktake Report does identify the risk of algorithmic harms, it also states:

(a) there is "no consistent approach to capturing and considering the views of key stakeholders during the algorithm development process";[147]

(b) following the deployment of an algorithm, there is "little consistency across government in formally undertaking" regular reviews and improvements "to ensure there are no unfair, biased or discriminatory outcomes", and that this presents an "opportunity" to implement formal safeguards;[148]

(c) agencies could benefit from sharing expertise (implying variable capability across government), including through the establishment of a "centre of excellence";[149]

(d) more could be done to "clearly explain how significant decisions are informed by algorithms";[150] and

---

145 Statistics New Zealand and Department of Internal Affairs, above n 2. See also Frith Tweedie "CIO Upfront: She'll Be Right? Government Review of Algorithms Shows Need for Caution" *CIO New Zealand* (3 December 2018).

146 The report notes that where there are automated processes, they are "usually restricted to decisions in favour of an applicant or client except where there is a clear degree of legal transparency related to automatic decision-making". Statistics New Zealand and Department of Internal Affairs, above n 2, at 30.

147 At 33.

148 At 29, 32 - 35.

149 At 32 - 35.

150 At 34.

(e) consideration should be given to ways to ensure stakeholder views are incorporated in the algorithm development process, and in particular a te ao Māori perspective, and to ensure privacy, ethics and human rights are considered as part of algorithm development and procurement, potentially through some form of impact assessment.[151]

Even if the Stocktake Report leads to better practice across the public sector, Liddicoat et al highlight how the Court Matters Act 2018 ("**CMA**") is possibly the only piece of legislation to date that expressly considers and regulates automated processing in New Zealand.[152] The CMA amends the Summary Proceedings Act 1957 by allowing the Ministry of Justice ("**MOJ**") to set up an automated system that approves or rejects applications to extend the time for payment of fines owed to MOJ or to vary payment arrangements.[153] While MOJ must create procedures to set the criteria for approval or rejection, and identify what information will be relevant, the applicant only needs to be made aware that the system's decision can be reviewed by a person[154] – there is no explicit requirement to notify the applicant about how the system works. MOJ must however be satisfied that the system "has the capacity to do any actions required with reasonably reliability", and that a person who asks for human review will receive this "without undue delay".[155] There is no requirement to consider any mitigations for the potential harms described above.

Other than the CMA, PSA use of algorithms is likely to be governed by existing statutory frameworks (such as the Privacy Act) and best practice guidelines which may constitute relevant, but not necessarily mandatory, considerations for administrative law purposes. Despite the Privacy Commissioner's suggestion that New Zealanders should receive a right to algorithmic transparency similar to that contained in the GDPR (discussed further below),[156] this been left out of the Privacy Bill that was recently reported back from select

---

151 At 33 - 34.

152 Joy Liddicoat, Colin Gavaghan, Alistair Knott, James Maclaurin and John Zerilli "The Use of Algorithms in the New Zealand Public Sector" (2019) NZLJ 26 at 29.

153 Court Matters Act 2018, s 61.

154 Summary Proceedings Act 1957, s 86DA.

155 Summary Proceedings Act 1957, s 86DC.

156 John Edwards "Privacy Commissioner's Submission on the Privacy Bill to the Justice and Electoral Select Committee" (31 May 2018) at 30 - 34. The Privacy Commissioner suggests that a gap remains in the existing legal protections: "the Privacy Act 1993, the Human Rights Act 1993, the New Zealand Bill of Rights Act 1990 and the Official Information Act 1982 provide the general legal human rights

committee.[157]

Recent reports have, however, drawn more attention to the possibility of regulation. Gavaghan et al recently considered the use of algorithms by the New Zealand Government.[158] They recommend (among other things) the establishment of an agency to provide best practice guidance and oversight, and a register of algorithms used in government.[159] At the same time, the AI Forum has argued for restraint until it becomes clear that regulation is needed.[160] The Human Rights Commission has also analysed the potential for privacy and human rights harms through the use of algorithms, without specifically calling for any further remedies or regulation.[161]

There are nevertheless several pieces of guidance that may influence PSAs' use of algorithms. The clearest example is the *Principles for Safe and Effective Use of Data and Analytics* developed by the Privacy Commissioner and Statistics New Zealand.[162] The *Principles* consist of a very short list of basic matters that an agency should consider when using algorithms: that they should deliver clear public benefit; that data should be fit for purpose; that a focus remains on people; that transparency is adequately maintained; that the tool's limitations are understood; and that human oversight is retained.[163] For its own purposes, MSD has also created a process-based framework in an attempt to avoid algorithmic harms, called the *Privacy, Human Rights and Ethics Framework* ("**PHRaE**").[164] Rather than providing a set of principles, the PHRaE creates a process for ongoing project design and feedback, and includes specialist support and the use of interactive tools.[165] The Social Investment Agency is also embarking on the creation of a

framework… but these Acts do not create any general and principled high level framework protecting individuals in relation to automated decision-making."

157 Privacy Bill 2018 (34-2).

158 Gavaghan et al, above n 81.

159 At 4. As we will see in chapter six, this thesis's ultimate recommendations for a regulatory modal broadly align with these suggestions.

160 See AI Forum New Zealand *Artificial Intelligence: Shaping a Future New Zealand* (March 2018) at 70.

161 New Zealand Human Rights Commission *Privacy, Data and Technology: Human Rights Challenges in the Digital Age* (Auckland, May 2018).

162 Office of the Privacy Commissioner and Statistics New Zealand *the Principles for Safe and Effective Use of Data and Analytics* (16 May 2018).

163 Ibid.

164 Ministry of Social Development *Privacy, Human Rights and Ethics Framework.*

165 Ibid.

draft "Data Protection and Use Policy" for its work,[166] although the extent to which this will focus on algorithmic hygiene is unclear. Agencies may also refer to the Data Futures Partnership's *A Path to Social Licence: Guidelines for Trusted Data Use* which asks agencies to consider eight questions focusing on "value", "choice" and "protection" in order to guide agencies towards actions which citizens will be comfortable with (rather than actions which are merely legally permitted).[167] Compliance with these frameworks may reduce the risks of harm and help to protect agencies from unlawful actions. However, the Stocktake Report indicates their influence has been limited in ensuring good practice.

Given these examples are voluntary and offer only modest, if any, real constraint on how most PSAs use algorithms, it is worth looking at steps towards regulation of algorithms in comparable jurisdictions.

2 Overseas regulation of algorithms

The risks of algorithms have increasingly grabbed the attention of academics, policymakers and civil society across the globe. In the United Kingdom, a series of Parliamentary reports have considered the legal and ethical challenges of big data,[168] algorithmic decision-making,[169] and artificial intelligence,[170] and in 2017 the United Kingdom Government announced the establishment of a Centre of Data Ethics and Innovation to advise on algorithmic and data issues.[171] Likewise, the United States' Obama Administration issued reports on the danger that big data[172] and algorithmic decision-making[173] perpetuates historical biases. The Australian Government has recently considered similar issues and sought feedback on an ethical framework to mitigate harms arising from the use of artificial

---

166 Social Investment Agency *What You Told Us: Findings of the 'Your Voice, Your Data, Your Say' Engagement on Social Wellbeing and the Protection and Use of Data* (November 2018).

167 Data Futures Partnership *A Path to Social Licence: Guidelines for Trusted Data Use* (August 2017).

168 House of Commons Science and Technology Committee *The Big Data Dilemma: Fourth Report of Session 2015-16* (10 February 2016).

169 House of Commons, above n 18.

170 House of Lords Select Committee on Artificial Intelligence *AI in the UK: Ready, Willing and Able?: Report of Session 2017 – 19* (16 April 2018).

171 United Kingdom Government "Centre of Data Ethics and Innovation" < https://www.gov.uk/government/groups/centre-for-data-ethics-and-innovation-cdei>

172 The White House *Seizing Opportunities*, above n 24.

173 The White House *Report on Algorithmic Systems*, above n 17.

intelligence across the public and private sector.[174] More generally, a base of literature has developed on the challenges of ethical use of algorithms and best practice.[175]

Despite an awareness of potential algorithmic harms, few countries have put in place significant additional protections. Canada has taken arguably the biggest step toward regulating PSA use of algorithms, with its *Directive on Automated Decision-Making*[176] recently issued under its Financial Administration Act.[177] The *Directive* is a secondary piece of legislation that does not create rights directly enforceable by citizens, but it does require agencies to work towards good algorithmic hygiene – for example, by requiring them to undertake an algorithmic impact assessment, imposing transparency and explainability requirements, and by requiring that appropriate quality assurance processes are built into the use of algorithms. As this thesis discuses in chapter six, aspects of the Canadian model are worth considering in New Zealand.

The GDPR,[178] which entered into force in May 2018, also provides some inspiration.[179] Broadly speaking, Art 22 of the GDPR gives a data subject the right not to be subject to a "decision based solely on automated processing" which produces "legal effects" concerning him or her. This right is subject to some exceptions, including consent and where a member state has provided safeguards to protect the subject's rights and freedoms. Processing should not occur where it concerns certain sensitive categories of data (designed to protect against discrimination),[180] except where there are reasons of "substantial public interest" or other exceptions apply (e.g., where explicit consent is obtained). The GDPR also requires data controllers (public or private) to undertake "data protection impact assessments" ("**DPIA**") if there is likely to be a "high risk to the rights and freedom of natural persons" including from automated or large scale processing of personal data, or

---

174 Australian Government *Artificial Intelligence: Australia's Ethics Framework – A Discussion Paper* (Canberra, April 2019).

175 See for example Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter and Luciano Floridi "The Ethics of Algorithms: Mapping the Debate" (July – December 2016) Big Data & Society 1.

176 Canadian Government, above n 3.

177 Financial Administration Act (RSC 1985, c, F-11), s 7.

178 Although it is subject to a range of important exceptions. GDPR, above n 4.

179 This builds on protections under its predecessor, the Directive 95/46/EC on the protection of individuals with regard to the processing of personal data [1995] OJ L281/31.

180 Art 9(1). These categories broadly align with the prohibited grounds of discrimination found in the HRA and discussed more in chapter four.

from processing sensitive categories of data.[181]

Elsewhere, in the United States a new Bill proposes safeguards to ensure private companies do not perpetuate biases in their use of algorithms, by requiring large businesses to undertake impact assessments in the vein of the GDPR. However, its passage is not assured and it does not focus on government decision-making.[182] New York City is, on the other hand, leading the way. It has set up a taskforce on automated decision-making to recommend a process to allow individuals to request information on automated decisions, ensure disclosure information about decision systems, and to suggest remedies for harm where protected categories of persons are disproportionately impacted.[183]

In 2007, the Australian Government issued non-binding guidance on using automated decision-making, and the risks under administrative law, but this is now dated.[184] More recently, the Australian Office of Information Commissioner has provided general guidance on data analytics under the Australian Privacy Act 1988, but this operates within existing rules for management of personal information.[185] While a proposed ethical framework for using artificial intelligence has been released for discussion, no concrete steps have been taken towards regulation.[186]

In summary, many comparable jurisdictions are only cautiously taking first steps towards regulation of algorithms used for government decision-making. However, those examples which do exist, such as the Canadian *Directive*, are broadly consistent with the regulatory model proposed in chapter six.

F The rise of algorithmic decision-making: conclusion

This section has shown what will be common knowledge to many inside governments:

---

181 Art 35.

182 Cyrus Farivar "New Bill Aims to Stamp out Bias in Algorithms Used by Companies" NBC News (11 April 2019).

183 New York City Automated Decision Systems Task Force "About" < https://www1.nyc.gov/site/adstaskforce/about/about-ads.page>

184 Australian Government *Automated Assistance in Administrative Decision-Making* (February 2007).

185 Office of the Australian Information Commissioner *Guide to Data Analytics and the Australian Privacy Principles* (March 2018).

186 Australian Government, above n 174.

public actors are increasingly using algorithms across the gamut of social services to help make decisions. In doing so, citizens face the prospect of real benefits, such as the amelioration of common biases and mental shortcomings, cost reductions and quicker services. At the same time, if we are not careful, a range of harms are possible: stretching from feedback loops which entrench existing biases, to failures in accountability, to cumulative harms arising from ongoing categorisation.

We have also seen that New Zealanders' consciousness of algorithmic decision-making is gradually becoming more apparent, as high profile examples – in areas such as immigration and the care and protection of children – hit the headlines. However, for now New Zealand has taken only baby steps towards confronting the potential for algorithmic harm. Despite deficiencies in New Zealand agencies' use of algorithms highlighted in the Stocktake Report, no formal steps towards regulation have been proposed and only soft guidance and existing legal frameworks will guide PSAs' use of algorithms.

There are however some learnings to be taken from overseas developments; in particular Canada's *Directive* and the EU's GDPR. We will return to how these frameworks can inform a regulatory model for New Zealand, in chapter six.

However, for now, the next three chapters explore the adequacy and extent of existing legal protections for individuals who may be affected by algorithmic decisions.

III Chapter Three: Informational Rights

A Overview

This chapter considers both the ways in which individuals' informational rights actually respond to the potential issues arising from PSAs' use of algorithms discussed in chapter two, as well as their potential to do so. Relevant informational rights discussed include those arising under the Privacy Act, OIA and LGOIMA, and s 21 of NZBORA.

This chapter will show how all three of these legal avenues, in principle, say something about the harms commonly associated with the use of algorithms. In particular, with appropriate tweaking, the Privacy Act could provide a useful remedial avenue in cases where individuals suffer real harm through algorithmic decisions. Moreover, this chapter suggests existing rights of transparency under the OIA and LGOIMA can be significantly expanded to enhance persons' ability to know about, understand, and seek human review of decisions about them. Lastly, the right to be free from unreasonable search and seizure under s 21 of NZBORA will provide limited remedies where algorithms are used for enforcement purposes, especially if courts adopt a more nuanced approach to expectations of privacy. However, this chapter also outlines how these changes are not a panacea for broader issues which affect the efficacy of these informational rights.

B Right for information to be kept accurate and complete

1 The right to accuracy

IPP 8 is likely to provide the Privacy Act's most useful ground of complaint for those affected by an algorithmic decision, even if it will not work in every case. IPP 8 provides that an agency will not use personal information:[187]

> … without taking such steps (if any) as are, in the circumstances, reasonable to ensure that, having regard to the purpose for which the information is proposed to be used, the information is accurate, up to date, complete, relevant, and not misleading.

In order to take legal action under IPP 8 of the Privacy Act, an individual will in most cases need to show an "interference with privacy" – effectively that objective harm has arisen

---

187 Section 6.

from a breach of an IPP.[188]

There are at least two ways in which IPP 8 could be relevant to an algorithmic decision. First, the PSA could fail to take reasonable steps to ensure the accuracy of information about an individual that it uses as *input data* for an algorithm. The application of IPP 8 in these scenarios is similar to any other (non-algorithmic) situation where a PSA has relied upon incorrect data to make a decision. If there has been a failure to ensure reasonable processes to protect against these mistakes arising, then an individual may be able to establish a breach of the IPP and bring a claim in the HRRT if he or she fails to reach settlement with the PSA. In this sense, IPP 8 provides a strong mechanism for complainants where there have been sloppy processes for ensuring the probity of primary data used in decision-making.

This first ground attacks a primary criticism of the use of algorithms – that is, how often they incorporate input data which is biased or simply wrong. One can imagine the many ways in which data errors can arise through the bureaucracy of the state: from simple recording errors (such as Work and Income incorrectly recording information provided by a client); to subjective matters of judgement that may reflect bias (e.g., police officers' ideas as to what is "suspicious"); to actual data entry errors (e.g., data being entered into the wrong field, or someone else's data being mistakenly used); to a simple failure to update data to reflect new circumstances (that is, relying on old information which is now invalid). In these numerous ways, individuals – to the extent they can discover the nature of the inputs used – may be able to assert a breach of IPP 8 when inaccurate data is used without reasonable precautions.

On the other hand, deliberate use of data that appears relevant but which is known to be potentially inaccurate will not necessarily be a breach of IPP 8 if the decision-maker has weighted this risk accordingly. Therefore, where an algorithm uses information known to have dubious reliability, the PSA will need to be able to demonstrate that the algorithm's use of this information was significantly discounted to account for this risk. However, without this evidence an affected individual may have good grounds for a claim.

The second way IPP 8 could be relevant is more particular to the use of algorithms. An individual may be able to argue that a PSA has failed to take reasonable steps to ensure the

---

188 Section 66 usually requires both the breach of an IPP and objective actual or likely loss or harm, which can include significant humiliation or loss of dignity.

accuracy of an algorithm's *output data* (for example, inferences, recommendations or predictions about an individual used in decision-making). This could arise because of a failure to ensure that the algorithm model has been validated to give the right answer to the right question (or if it fails to do so when applied to outlier groups), because the algorithm inherently has a low standard of accuracy. This could also arise because (particularly in the case of ML algorithms) a lack of oversight has created a self-justifying feedback loop.

2 Outputs as personal information?

This second scenario raises an immediate question: is algorithmically generated information relating to an individual that individual's "personal information" (being information about an identifiable individual) under the Privacy Act? An intuitive reaction might be that – given the same kind of harms can arise – this "inferred" output data should be, but there remains uncertainty as to whether this is the case in New Zealand. Nevertheless there is support for this position from at least two overseas authorities.

First, the Office of the Australian Information Commissioner ("**OIAC**") suggests that inferred data is "personal information" under Australia's very similar, principles-based Australian Privacy Act 1988 (*Cth*):[189]

> Data analytics can lead to the creation of personal information. For example, this can occur when an entity *analyses a large variety of non-identifying information, and in the process of analysing the information it becomes identified* or reasonably identifiable. Similarly, *insights about an identified individual from data analytics may lead to the collection of new categories of personal informatio*n…
>
> (emphasis added)

This position is also supported by *Ewert v Canada*, mentioned above.[190] There, the Supreme Court of Canada considered that a statutory obligation to "take all reasonable steps to ensure that any information about an offender that it uses is as accurate, up to date and complete as possible" applied to information created by actuarial risk prediction tools used by the Correctional Service of Canada.[191] While that finding turned partly on the

---

189 Office of Australian Information Commissioner, above n 185, at 23.

190 McGovern explores how IPP 8 might apply in a *Ewert*-type situation if it occurred in New Zealand. See Danica McGovern "Ewert v Canada (2018) SCC 30" [2019] NZLJ 131.

191 *Ewert v Canada*, above n 22.

statutory context, the Court also considered that algorithmic results were "information about an offender" in the ordinary meaning of those words.[192] This wording closely mirrors the Privacy Act definition of personal information as "information about an identifiable individual".[193]

From a first principles basis, there are good arguments that inferred data should be an individual's personal information.[194] Arguably there is little difference between information about an individual generated by a machine and assigned to an individual, compared with similar information generated by a human. For example, it would be hard for an employer to argue that a yearly employee evaluation created with a set of standard questions and a scoring sheet is not the employee's personal information. It is hard to see a significant difference between this case, and the situation where a software algorithm carries out an evaluation according to a model created by a PSA. Additionally, there is a good argument for protecting inferred information to the extent that it provides clues (or can be reverse-engineered) to reveal primary data that is personal information.

Second, whether human or machine generated, this kind of output data can be relied upon for important decisions about individuals as if it is reliable information "about" them. Where the output is lacking in accuracy, this can lead to real harm of the kind IPP 8 aims to prevent. So, in the sentencing-bias scenario described above, a lack of oversight could lead to a false "risk" score being assigned to a person, affecting sentencing. Or, in the benefit surveillance scenario an individual could be put in the "wrong bucket", cutting off the individual's income.

Therefore, to ensure the Privacy Act can definitely respond to the policy goals which underlie IPP 8, the Act should be amended to explicitly include algorithmically-generated outputs as potentially within the definition of "personal information". Because the same harms can arise as with primary data, secondary outputs should be personal information at least to the extent they "say something" about an individual – for example, the person's likelihood of having particular characteristics (such as being vulnerable to family violence)

---

192 At [33].

193 Section 2.

194 Even if the individual is profiled into a broader group, which creates individuals who will have similar claims that the information is "their" personal information. See the discussion of the challenges of profiling for typically individual-focused data protection regimes by Edwards and Veale, above 42, at 35 - 36.

or doing something in future (such as offending).[195] Importantly, a range of exceptions within the Privacy Act would still allow this information to be shared and used by PSAs without necessarily needing an individual's "authorisation" (for example, where this is to avoid prejudice to the maintenance of the law).

3 The balance between accuracy and public policy choices

Reliance on IPP 8 in relation to inferred information raises a further question: if IPP 8 requires an agency to take reasonable steps to ensure the accuracy of personal information, what is reasonable for these purposes? Consider an algorithm which is, by design, known to be accurate 98 per cent of the time. Does this threshold mean an agency should be presumed to have taken reasonable steps? What about where the algorithm is *severely* wrong in those 2 per cent of cases or, because of how the algorithm is used, being in the unlucky few will have significant ramification? Even if the agency regularly audits its algorithm and takes steps to ensure the reliability of its input data, is that enough in these cases?

These questions highlight the challenge PSAs face to weigh public benefits and detriments arising from the use of algorithms, and the difficulty squaring this calculus within the individual-focused privacy rights provided by the Privacy Act. However, IPP 8's reference to reasonableness and the relevant circumstances also suggests that the acceptable degree of harm arising from false positives or negatives – depending on how the algorithm is tuned – is likely to depend on the nature of the algorithmic decision, and the potential adverse consequences for a person tarred with an inaccurate result. A high chance of a false positive or negative may be acceptable when the likely result is relatively trivial, or the decision-maker will significantly discount his or her level of reliance given known potential inaccuracy (as outlined above). However, in some other situations – such as where an offender is given a disproportionate risk rating, materially influencing a judge to increase her sentence from home detention to a period of imprisonment – the consequences can be more significant. In those cases, an individual may fairly be able to say: "put aside the general accuracy of the algorithm – in *my case* what did you do to ensure that the output was accurate?"

Consider this question in relation to the previously discussed MSD algorithm designed to

---

[195] Gavaghan et al have also queried whether the Privacy Act should be expanded to include inferred data. See Gavaghan et al, above n 81, at 76.

detect vulnerable children. This has a 37 per cent accuracy of detecting children likely to suffer harm in the first five years of their life. Given the total population rate of children who suffer harm is 5.4 per cent, the algorithm provides the prospect of significant harm reduction for children who otherwise might not be reached.[196] However, its use also means that almost two thirds of the caregivers of such children will be inaccurately classified as potential child abusers and may be subject to unwarranted "hypervigilance" from state agencies.[197] These individuals may face the stigma of increased social service intervention for suggested *future* events which cannot be verified, and harm through "false accusations or the incorrect removal of children".[198] These interventions could themselves create path dependency toward other undesirable outcomes which increase the risk of harm to children (e.g., avoidance of engagement with state agencies). While an ethical review has suggested that, on balance, this harm is offset by the expected benefits to vulnerable children,[199] that may not matter for the purpose of the Privacy Act. It seems likely that those affected could still argue that the PSA knowingly relied on information that was more likely than not to be inaccurate and harmful to them, leading to an interference with their privacy.

To address this issue, the application of IPP 8 to algorithmic use cases should also be clarified as part of the regulatory model described in chapter six. While there is a good argument IPP 8 should require reasonable steps are taken to ensure accuracy, both for the whole population to which the algorithm will apply, and to any reasonably discernible sub-groups, this approach may be too broad brush given the state's role in making difficult policy trade-offs. For example, it may be appropriate to use an algorithmic tool that could cause harm through a high number of false positives, if the consequence of any false negatives is particularly severe. In this case, the state may choose not to take "reasonable steps" to ensure accuracy because it knows that accuracy (in the sense of improperly classifying people) is unattainable to achieve the relevant policy goal of few false negatives. To address this issue, this thesis suggests a regulatory framework which creates political accountability for *intentional* policy trade-offs, while inoculating PSAs from legal challenge in these cases. This is discussed in more detail in chapter six.

---

196 Rhema Vaithianathan et al, above n 139, at 3.

197 Anton Blank, Fiona Cram, Tim Dare, Irene De Haan, Barry Smith and Rhema Vaithianathan *Ethical Issues for Māori in Predictive Risk Modelling to Identify New-Born Children who are at Risk of Future Maltreatment* (January 2015) at 7.

198 At 8.

199 Dare, above n 140, at 25 - 32.

In summary, IPP 8 provides a possible protection against algorithmic harms – particularly if changes clarify outputs can be personal information. However, it will not always respond to harms (e.g., if an agency *has* taken reasonable steps, but errors nevertheless occurred in a particular case), while the regulatory model proposed in chapter six could legitimise inaccuracy in clear situations involving difficult policy trade-offs.

C Rights relating to collection of information

1 Right to receive notice

A right for individuals to receive notice when algorithmic outputs are created would respond to concerns about transparency of decision-making, outlined above. However, as currently interpreted, real doubt remains that IPP 3 of the Act would automatically create this obligation.

Where an agency collects personal information directly from an individual, IPP 3 requires the agency to take reasonable steps to ensure the individual is made aware of specified information about the collection. This includes the fact and purposes of collection, any intended recipients of the information, the name and address of the agency, whether collection is required or authorised by law, the consequence of the individual not providing information, and the individual's rights of correction and access to the information under the Privacy Act.[200]

If the outputs from an algorithm can be personal information, this raises a question: has there been direct "collection" such that IPP 3 requires the relevant individual to be notified? If IPP 3 does require this notice, and harm arises in connection with this failure to notify, this would create grounds for a person affected by an algorithm's output to claim an "interference with privacy" and the ability to seek remedies under the Privacy Act.

In principle, agencies should have to tell to individuals when they create new algorithmic information to make decisions about them. Otherwise individuals cannot easily contest or ensure accountability for how it is used. Moreover, the structure of the Privacy Act supports the proposition that when this information is created it is "directly" collected by the PSA . IPP 2 sets up a range of exceptions to the presumption that information should be collected *directly*. Implicitly these exceptions suggest that when an agency collects information

---

200 Section 6.

indirectly, this is because someone has already collected it "directly" (i.e., indirect collection is necessarily subsequent to another party's direct collection). In general, the exceptions in IPP 2(2) apply where an agency needs to get the information from someone already holding the information or where it is accessed from elsewhere (e.g., because it is publicly available). Because these situations do not apply where a PSA's algorithm creates new information, this implies a PSA must be the first "direct" collector of the information.

However, a court may well disagree. First, given the Privacy Act applies to both public and private agencies, there is an argument it is unworkable to have a proactive requirement to give notice to individuals every time new information is inferred about an about an individual.[201] Second, the natural meaning of "collect" implies the taking from another (rather than the creation from existing information). Third, as Roth argues, based on the Privacy Act's modelling on the OECD Guidelines,[202] "directly collect" was intended to relate to circumstances where the individual would be aware that collection was taking place – which is obviously different from algorithmically inferred information.[203] Arguably, in other situations where an agency comes into information indirectly (e.g., information an employer acquires about an employee from others), this information is "obtained" not collected (and so no notice obligation arises).[204] On the other hand, even if this position reflects the historical intent, this should not prevent a court from interpreting the Act to accommodate a world where inferred information is regularly created and used in relation to individuals.

Given doubt about whether inferred information is "collected" by an agency when it is created, the Act should be amended to expressly provide for this. Importantly, such an amendment would ensure the constraints of IPPs 1 and 4 would apply to protect against inappropriate use of the information (discussed more below). Although the workability objection is legitimate, changes could ensure that, for private entities, the obligation to notify would only apply subject to a materiality threshold based on the potential impact on the individual (and could be subject to the exceptions already found IPPs 2 and 3). However, where the information was created by a PSA, this thesis suggests individuals

---

201 Although it should be noted that most private agencies will provide adequate notice by having an up to date privacy policy that describes this process.

202 See the Explanatory Memorandum to the Organisation for Economic Co-operation and Development *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* (1980).

203 Paul Roth *Privacy Law and Practice* (online looseleaf ed, LexisNexis NZ Limited) at PVA 6.6.

204 At PVA2.5.

could instead rely on a separate right to know when algorithms are being used by the state to make meaningful decisions about them (discussed more below in relation to the OIA and LGOIMA).

2 Right against unfair and unlawful collection: the "passive" surveillance problem

Next, we will examine how IPP 4 of the Privacy Act shows promise for addressing unfair forms of surveillance by PSAs, but is hamstrung by a constrained interpretation of "collect".

IPP 4 requires an agency not to collect information by means that are unlawful or which, in the circumstances, are unfair or unreasonably intrude upon an individual's personal affairs.[205] In principle, IPP 4 provides an obvious bulwark against the unfair use of algorithmic surveillance tools such as AFR, gait recognition technology, licence plate scanners and social media surveillance tools.

However, it is unclear whether IPP 4 applies to AFR and similar tools to the extent that they are "passive" collectors of information. Although the Privacy Commissioner and HRRT have implied one can unfairly collect information through things like hidden video cameras,[206] the Court of Appeal's judgment in *Harder v Proceedings Commissioner* remains precedent for the view that information is not "collected" where it is passively recorded, because it is not "solicited".[207] Roth also argues this position accords with the legislative intent.[208] While the HRRT[209] and the Law Commission[210] have attempted to walk this position back to its narrowest point, the broader position remains unsettled until a superior court clarifies the scope of IPP 4. Without the protection of IPP 4, the right to bring a claim might only exist for very intrusive surveillance that meets the high test for the tort of intrusion into seclusion.[211]

Supposing the Privacy Act does apply to collection of information via surveillance, the

---

205 Section 6.

206 Roth, above n 203, at PVA2.5.

207 *Harder v Proceedings Commissioner* [2000] 3 NZLR 80 (CA).

208 Roth, above n 203, PVA6.6.

209 *Armfield v Naughton* [2014] NZHRRT 48.

210 See Law Commission *Review of the Privacy Act 1993: Review of the Law of Privacy Stage 4* (NZLC R123, 2011) at [2.81].

211 See *C v Holland* [2012] NZHC 2155, [2012] 3 NZLR 672.

standard under IPP 4 is also high. For example, the Privacy Commissioner did not consider IPP 4 was breached when MSD required non-government social service providers (who relied on government funding) to collect client-level information in order for individuals to access their services.[212]

Arguably IPP 4 also focuses on the wrong area; individuals will usually be more concerned with how information is used, rather than whether it was unfairly collected. Consider AFR used in public spaces or online. The objection with this technology is not necessarily the collection of the information itself but, rather, how facial imagery is used in entirely new ways: an automated check to verify whether a person has previously offended and hence could be a "risk",[213] or the ability to determine by the shape of an individual's face whether they are more likely to be straight or gay.[214]

Despite these objections, the Privacy Act's definition of "collect" should be changed to override *Harder* so that IPP 4 can provide at least some measure of protection against unfair surveillance-type activities. This would ensure the Privacy Act can respond to current needs without being held back by definitional technicalities. In those cases which meet the high threshold of unfairness, it is appropriate that individuals should be able to claim an interference with privacy and to access remedies.

D Rights of access and transparency

1 Rights to information under the OIA and LGOIMA, and Privacy Act

Separate from rights relating to collection, the Privacy Act, OIA and LGOIMA create rights which respond to concerns about the transparency of algorithmic decisions, potentially helping individuals challenge these decisions. However, the following discussion illustrates how this position would be substantially improved through more significant rights to proactive notice of algorithmic decisions, a right to human review, and further clarity about how a right to reasons under the OIA and LGOIMA applies to algorithms.

---

212 While acknowledging that this might, however, not be the case for some aggrieved individuals. See Privacy Commissioner, above n 138, at 31.

213 See Georgetown Law Centre, above n 21, at 330, 365 - 369, and 377.

214 Sam Levin "New AI Can Guess Whether You're Gay or Straight from a Photograph" *The Guardian* (8 September 2017).

The OIA provides general rights to access "official information"[215] held by most PSAs, and also gives persons a right to access reasons for the decisions made about them. This right to reasons provides the right, subject to some withholding grounds, to request and be given a written statement of: (a) the finding of material issues of fact; (b) a reference to the information on which the findings were based; and (c) the reasons for the decision or recommendation.[216] The LGOIMA provides almost identical rights in respect of local government entities[217] and so, for convenience, the remaining references in this thesis to the OIA should also be read to include the LGOIMA.

IPP 6 of the Privacy Act meanwhile allows individuals to request their personal information held by an agency. In principle an individual can request any information about them used in a decision, including any relevant input data and potentially inferred output data (depending on the interpretation discussed above). To the extent a request under the OIA relates to personal information, that element of the request is considered under the Privacy Act rather than the OIA.[218]

If a PSA fails to comply with the access rights under the OIA and Privacy Act, the affected party has potentially significant recourse. Under the OIA, a failure to give adequate reasons may provide grounds for judicial review of the PSA (discussed in chapter five). This leaves open to what extent a PSA will be able to satisfy the requirements of the OIA in circumstances where it cannot provided a detailed description of how the algorithm reaches its decision (let alone the software source code). As outlined in chapter two, this could be particularly challenging in the case of ML algorithms which cannot be interrogated by humans.

Likewise, where a PSA is unable to provide an individual with his or her personal information in accordance with the Privacy Act, this will automatically constitute an "interference with privacy" allowing the individual to take action. Moreover, unlike the OIA, when a request reveals personal information that has, or will be used, by an algorithm and this information is incorrect or inaccurate, IPP 7 of the Privacy Act provides a process whereby the individual can seek correction of that information. Depending on whether inferred information is "personal information" this could provide a strong back-door

---

215 Section 12.

216 Section 23.

217 Local Government and Official Information and Meetings Act 1987 [LGOIMA], s 22.

218 OIA, s 12(1A); LGOIMA, s 10(1A).

challenge to the information relied upon in a decision.

2 Rights to information: issues and obstacles

However, there are a number of issues which limit the effectiveness of IPP 6 and/or the OIA as an effective means of guaranteeing access to information. First, the obligation is necessarily reactive – there is no obligation on a PSA making a decision to notify affected parties. This severely limits the utility of these rights to respond to the challenges of algorithms. Individuals will not know to ask for information.

Second, even where a person does ask, both the OIA and Privacy Act allow agencies to extend timeframes for their responses beyond a presumptive 20 working days, despite an obligation to provide the response "as soon as reasonably practicable".[219] Agencies can also transfer requests to other agencies,[220] grant themselves extensions where a large quantity of information is involved or consultations are necessary to ensure a proper response,[221] and (in respect of the OIA) can charge fees depending on the costs of labour and materials to respond.[222] These powers can slow the process and disincentivise access. As Price has observed of the OIA, simple requests are often dealt with quickly and efficiently, but more complicated requests can founder and drag, sometimes for years.[223] A cynical view is that, depending on the scope and sensitivity of the information, requests about an agency's use of an algorithm could fall into the second camp. An unwillingness to respond to analogous freedom of information requests has certainly been observed in comparable jurisdictions overseas.[224]

Third, both the Privacy Act and OIA provide withholding grounds that can limit the information made available to a requester. The withholding grounds across these pieces of legislation are largely the same. However, while a PSA responding to an IPP 6 request can

---

219 OIA, s 15; Privacy Act 1993, s 40.

220 OIA, s 14; Privacy Act 1993, s 39.

221 OIA, s 15(a); Privacy Act 1993, s 41.

222 OIA, s 15(2).

223 Steven Price "The Official Information Act: Does it Work?" (2016) NZLJ 276 at 276.

224 For example, the United Kingdom's Department of Work and Pensions has refused to provide information about data inputs being used for new algorithmic tools being developed for public welfare services, including a tool to determine the "likelihood that citizens' claims about their childcare and housing costs are true when they apply for benefits". See Robert Booth "Benefits System Automation Could Plunge Claimants Deeper into Poverty" *The Guardian* (14 October 2019).

simply withhold information if it believes a relevant test is met, for most applicable grounds under the OIA the PSA will need to balance the ground against the public interest in and presumption towards disclosure.[225] Particularly relevant is a ground allowing withholding where the information would disclose a trade secret or unreasonably prejudice the commercial position of a party[226] – potentially limiting the ability to hold accountable a PSA that has engaged a commercial algorithms provider. However, this ground is subject to a public interest balancing test under both the OIA and (untypically) the Privacy Act, meaning that the broader interest in disclosure may prevail.

Fourth, the right to reasons under the OIA is couched in reasonably limited terms. For example, an agency need only make "reference" to the information on which the findings were based. Moreover, Taylor suggests the obligation to provide reasons is likely to be lower than anything which might be expected of a judicial officer,[227] meaning that a detailed explanation of the algorithm and the data used may not be necessary (but this is circumstance-dependant).

This thesis therefore suggests three key changes to the OIA to address the issues described above, and help provide public confidence in PSAs' use of algorithms: first, a right to notice that an algorithm is being used for a decision; second, a right to human peer review of an algorithmic decision; and third, clarification of how the right to reasons under the OIA applies to algorithmic decisions. These changes would supplement the regulatory model described in chapter six, and are consistent with the goal of transparent public administration underlying the OIA.

3 A new right to notice

First, there should be a proactive obligation for PSAs to notify individuals when algorithms are used to make decisions about them, as alluded to in the earlier discussion regarding notice under the Privacy Act. This obligation needs to be workable. Agencies should not have to notify for every trivial decision in which an algorithmic tool (widely construed) is used (e.g., use of Microsoft Word). This means there should be a materiality threshold based on the degree of reliance. This threshold should be set relatively low, as the PSA

---

225 OIA, ss 4, 5 and 9.

226 OIA, s 9(2)(b); Privacy Act 1993, s 28.

227 Graham Taylor *Judicial Review: A New Zealand Perspective* (4th ed, LexisNexis NZ Limited, Wellington, 2018) at 326.

would only have to notify the individual that an algorithm is involved in decision-making, that a right to reasons is available under the OIA, and that in certain cases a right to human peer review may also exist (discussed below). The PSA might also need to provide information about how the algorithm is used, but this would be on a proportional basis based on the use case (as outlined in later discussion on the regulatory model). Similar to how IPP 3 of the Privacy Act requires reasonable steps in the circumstances to provide notice of matters regarding collection of information, this PSA obligation could be tempered by what is reasonable in the circumstances.

4 A new right to human review

Alongside a right to notification, individuals should have a complementary right to have an algorithmic decision reviewed by a human. This should apply to decisions where the algorithm is substantially relied upon in the decision, and where the decision itself can reasonably be expected to have a material impact on the rights, entitlements, benefits or privileges of the individual. This right could be subject to any legislation that specifically provides for an alternative review or appeal process in relation to the particular decision, or which excludes this right.

The requirement for "substantial" reliance is intended to overcome criticism of the GDPR's right to object to *solely* automated decisions. While many decisions are informed by an algorithm, few are completely automated.[228] And yet decision-makers can still be subject to significant automation bias where there is substantial input from an algorithmic tool, meaning a right to review would be emaciated if limited solely to automated processing. As Edwards and Veale also point out, a more limited right would not address well-known algorithmic "war stories", such as the use of racially biased predictive tools in criminal justice sentencing.[229] This change therefore would be broad enough to catch "rubber stamping" of algorithmic recommendations.[230]

The second element (material impact) ensures the right to review is not hijacked by vexatious claims. It also bears resemblance to the GDPR standard giving rise to the right to object (where processing "produces legal effects" or "significantly affects" the

---

228 Edwards and Veale, above n 42, at 44 - 48.

229 At 45.

230 House of Commons, above n 18, at 34.

individual).[231]

However, we should not assume a right to peer review always assures a "better" outcome. As outlined in chapter two, humans are subject to a range of biases and mental shortcomings, and we may overestimate the value of a machine's output. There is also a danger that having a "human in the loop" simply creates a placebo effect – making us feel that potential risks have been addressed without proof that this is the case.[232] There is also an ethical danger if we impose a double standard where an algorithm can be shown to significantly outperform the accuracy of humans,[233] because it suggests the human decision may be more likely to cause harm than simply relying on the algorithm.

However, unpacking the differences between how algorithms and humans make decisions provides support for a right to human review. First, algorithms often fail to fulfil the normative expectations of human decision-making. Scantamburlo et al argue that when human decision-makers make decisions:[234]

> … they do so in accordance with a commonly understood set of normative principles, for example:
>
> - justice (e.g. equality before the law, due process, impartiality, fairness);
> - lawfulness (e.g., compliance with legal rules);
> - protection of rights (e.g., freedom of expression, the right to privacy).
>
> … the current generation of intelligent algorithms make decisions based on rules learnt from examples, rather than explicit programming, and often these rules are not readily interpretable by humans. There is thus no guarantee that intelligent algorithms will necessarily internalise accurately, or apply effectively, the relevant normative principles; nor that the system within which they are embedded will have the means to facilitate the meaningful exercise of particular normative obligations

The likelihood of these normative principles failing to be captured by an algorithm itself

---

231 GDPR, Art 22.

232 Gavaghan et al, above n 81, at 75.

233 See John Zerilli, Alistair Knott, James Maclaurin and Colin Gavaghan "Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?" (September 2018) *Philosophy & Technology* 1.

234 Teresa Scantamburlo, Andrew Charlesworth and Nello Cristianini "Machine Decisions and Human Consequences" (2018) (draft chapter for the forthcoming book K Yeung and M Lodge (eds) *Algorithmic Regulation*) at 3.

justifies the right to review. It also highlights the tension between "fairness" in decision-making understood as *correctness* or as *equity*, described in chapter two. But humans – unlike algorithms – tend to apply "moral sensitivity" to the correctness of their actions.[235] Algorithms can have better accuracy as against a prescribed criteria, but accuracy is not enough alone when algorithms operate on the basis of correlation and cannot see when the exception to the rule is needed.[236]

Moreover, a right to human review also protects confidence in public decision-making. It protects citizens' rights to be heard by the state's elected and/or accountable officials in accordance with natural justice – reflecting the normative principles above. Importantly, it also protects the *perception* of fairness in public decision-making. A recent study of managerial decision-making algorithms highlights that while people do not mind algorithms making "mechanical" type decisions (e.g., allocating work between people), they are far more likely to distrust, consider unfair and feel negative towards an algorithm's decision-making where it relates to a typically "human" type decision, such as staff hiring or evaluation.[237] In these cases, people may perceive that algorithms lack intuition or the ability to evaluate social interaction, or that being judged by a machine is itself demeaning.[238] In at least some applications of algorithmic decision-making by PSAs we might expect similar reactions, and the right to human review will be an important antidote.

Last, adopting a soft form of the "precautionary principle"[239] popularised in the environmental context, the unknown harms from algorithms should encourage a slow departure from the longstanding status quo of human decision-making. Chapter six explores in more detail why this approach is important.

It is worth acknowledging that practical steps can be taken to limit how an initial decision influences the reviewer. Gavaghan et al propose a smart suggestion: require the human reviewer to look at the decision *de novo* without information about what the algorithm recommended.[240] The two results could be then be cross-checked, with the reviewer able

---

235 Gandy, above n 66, at 35.

236 Edwards and Veale, above n 42, at 28.

237 Lee, above n 127, at 11 - 12.

238 At 12.

239 See David Kriebel, Joel Tickner, Paul Epstein, John Lemons, Richard Levins, Edward L Loechler, Margaret Quinn, Ruthann Rudel, Ted Schettler and Michael Stoto "The Precautionary Principle in Environmental Science" (2002) 109 Environmental Health Perspectives 871.

240 Gavaghan et al, above n 81, at 74 - 75.

to ascertain any clear outliers. In the meantime, further research should be undertaken into the effectiveness of human peer review and the circumstances which are likely to provide the most reliable outcomes.

5 An expanded right to reasons

Additionally, the right to reasons under the OIA should be amended so that individuals receive meaningful information about how an algorithm has contributed to a decision about them. Arguably transparency should be the "default" position when algorithms affect the public; when these effects could be significant, there should be a combination of explanation and as much transparency as possible.[241] However, it is unclear how the right to reasons does or should apply to algorithms, and in particular ML algorithms.

The OIA could provide a broad requirement that an agency must, where requested, provide an explanation of: (a) the logic or reasoning behind the algorithmic model in general; and (b) how the algorithmic model contributed to the particular decision at hand. This should be commensurate with the potential impact on the individual. For example, Babuta et al suggest that in relation to criminal justice applications, "it should be possible to retroactively deconstruct the model… in order to identify which factors led to a certain decision being made".[242] This reflects the high stakes nature of criminal justice decisions. In other cases, this requirement would be disproportionate.

The Algorithms Watchdog – the new proposed regulator discussed later – could provide guidance on what is sufficient for different cases. Canada's *Directive* provides a good starting point, creating four categories of response for both notice and explanation, depending on the potential impact of the decision.[243] So, for a decision which has a very low impact, plain-language "frequently asked questions" about an algorithm's processes, supplemented by a general non-technical description of the reasoning for the decision at hand, might be sufficient. However, for more high stakes decisions, reasons could require: (a) a more detailed technical description of how the model works; (b) the nature of the training data used; (c) how the model has been validated and any relevant audits or reviews that have been undertaken; and (d) information about how the model contributed to the decision at hand, ideally via a retroactive deconstruction of this through "logging" of the

---

241 House of Commons, above n 18, at 4.

242 Babuta et al, above n 22, at 19.

243 Canadian Government, above n 3, Appendix C – Impact Level Requirements.

decision. Following consultation, the Algorithm Watchdog's guidance could also indicate what is required for certain ML algorithms – such as neural networks – which are hard to interrogate, and whether their lack of technical transparency disqualifies their use in high stakes cases.[244]

This right to reasons could be subject to appropriate exceptions – for example, to avoid the possibility of "gaming" of the algorithm, or where due to the sensitive nature of the area (e.g., national security) it is inappropriate for the information to be disclosed. In other areas, such as criminal justice, it may be appropriate to require PSAs and/or their algorithm providers to provide expert witnesses in proceedings which concern the use of an algorithm.[245]

With rights to receive notice and access human review, and a proportional right to reasons, the OIA would provide a much improved framework to help protect those subject to algorithmic decisions.

E Rights to have data use confined to specific purposes

Next, it is worth considering the extent to which the Privacy Act's limitations on the use of personal information could have something to say about algorithmic decisions. In general, IPPs 1 and 10 suggest parameters preventing overly broad use (or re-use) of information. However, these use parameters have been interpreted permissively and a range of exceptions can apply to permit other new uses; this makes IPPs 1 and 10 each less reliable avenues for individuals to challenge an algorithmic decision.

1 Purpose limitation: not so limited

IPP 1 requires that agencies only collect information that they need – that is, an agency cannot collect personal information unless the information is for a "lawful purpose

---

244 Different levels of transparency are possible based on the kind of machine learning algorithm. A "Random Forrest" algorithm which is "supervised" can be retroactively deconstructed and assessed; an unsupervised neural network cannot, although "indirect inference" might be possible. See Babuta et al, above n 22, at 18 - 19.

245 Babuta et al, above n 22, at 22. A right to an explanation may be of little value of itself in relation to criminal justice applications, unless other resources are also available to those accused. See the Law Society of England and Wales *Algorithms in the Criminal Justice System* (June 2019) at 5.

connected with function or activity of the agency" and the collection is "necessary for that purpose". Once collected, IPP 10 requires information "obtained in connection with one purpose" not be used "for any other purpose", unless one of a range of exceptions apply. IPP 9 meanwhile states that an agency is not to keep personal information "for longer than is required for the purposes for which the information may lawfully be used".

These requirements focus on "data minimisation" and stand at odds with data mining, which encourages large-scale collection and re-use of data. As the OAIC has noted in relation to the Australian equivalent to IPP 1:[246]

> This principle may appear to challenge the concept of using 'all the data' for 'unknown purposes'. However, just because data analytics *can* discover unexpected or 'interesting' correlations, this does not mean that the new personal information generated is necessary to the legitimate functions and activities.

Nevertheless what is "connected with a function or activity of an agency" has been interpreted loosely, despite on its face implying that information should only be used for things which an individual might reasonably expect. Similarly, Roth notes that the requirement that the information be "necessary" as "thus far developed in New Zealand is not particularly strict".[247] Necessary has been interpreted to mean *reasonably* necessary or as required for given situation, rather than the higher standard of indispensable or essential.[248]

Therefore, to the extent that an agency can assert that the use of personal information via an algorithm is within the broad range of functions the PSA might be expected to carry out, it is unlikely that IPP 1 will have much to say. A breach is not impossible though: MSD's requirement that social service providers compulsorily collect client-level information in connection with the "social investment approach" was considered by the Privacy Commissioner to be a breach of IPP 1 because the reasons for collection were not clear.[249]

---

246 Office of Australian Information Commissioner, above n 185, at 22.

247 Roth, above n 203, at PVA6.4(c).

248 See *Tan v New Zealand Police* [2016] NZHRRT 32 (18 October 2016); and *Lehmann v Canwest Radioworks Ltd* [2006] NZHRRT 35 (26 September 2006).

249 See Privacy Commissioner, above n 138, at 22.

2 Exceptions to use restrictions

Relatedly, IPP 10 provides limits around how an agency can use an individual's personal information, in principle providing protection where information collected by an agency for one purpose is then used for a different purpose as part of an algorithmic decision. However, a PSA could rely on a range of exceptions to this rule. For example, an agency does not have to comply with IPP 10 if the purposes are to "avoid prejudice to the maintenance of the law by a public sector agency" or "for the protection of the public revenue". This gives Inland Revenue and the Police a potentially wide-berth around the normal protections in the Privacy Act, as well any other agency using an algorithm to detect suspected fraud or take enforcement action.[250] IPP 10(1)(f)(ii) also provides a significant exception where the information is "used for statistical or research purposes and will not be published in a form that could reasonably be expected to identify the individual concerned". As an algorithmic tool might be read as falling within "statistical or research purposes", and the information will usually be used *internally* by a PSA rather than published, this exception potentially drives a hole through the normal data use protections. Moreover, the use limitations in IPP 10 are subject to any law that authorises or requires an agency to undertake a particular action.

Therefore, on current interpretations, IPPs 1 and 10 are unlikely to provide much assistance to someone subject to a harmful algorithmic decision. Moreover, given these principles apply to all uses of personal information, and across both public and private sector agencies, we should be cautious about suggesting any changes to the Privacy Act to address the above issues as they relate to algorithms. Instead, concerns about overly broad use and re-use – including the scope of the "research purposes" exception – are addressed by the regulatory model outlined in chapter six which legitimates PSA use of algorithms within clear boundaries.

F Rights against unreasonable search and seizure

Next, we consider the extent to which the right to be free from unreasonable search and seizure responds to potential algorithmic harms. Section 21 of NZBORA provides that everyone "has the right to be secure against unreasonable search or seizure, whether of the person, property, or correspondence or otherwise." As with other rights discussed in this

---

250 Section 6, IPP 10(1)(c).

chapter, the policy rationale for a right to be free from unreasonable search and seizure – to protect against unwarranted intrusions by the state into individuals' private affairs – would seem to respond to unfair uses of algorithms for surveillance. While use of surveillance algorithms, by themselves, will not always involve a decision of the kind typically addressed in this thesis (e.g., whether a person should be deported or not), the outputs of algorithmic surveillance can contribute to these decisions (e.g., if AFR falsely matches a person's face with a known overstayer, and action is taken on this basis).

However, as we will see, without further tweaks s 21 of NZBORA may not provide great assistance to those subject to algorithmic surveillance techniques.

1 The "enforcement" limitation

First, while there is some debate, it is likely that s 21 only applies to "enforcement" functions of the state. Historically the right has been limited to law enforcement functions, and possibly some forms of regulatory enforcement.[251] So, it may not apply where information gathered is used to make decisions about entitlements to state resources (e.g., access to a benefit). A lacuna would therefore exist if neither the right to be free from unreasonable search and seizure nor the Privacy Act applies to unfair collection of information recorded through "passive" surveillance devices or techniques.

2 Reasonable expectations of privacy: the public spaces issue

Moreover, it is unclear whether s 21 will prevent the widespread use of AFR by the state in "public" spaces (physical or online). While this technology can obviously be justified for relevant use cases, the Supreme Court's position in *Hamed v R* implies that individuals will generally have no expectation of privacy in public spaces:[252]

> If the surveillance is of a public place, it should generally not be regarded as a search (or a seizure, by capture of the image) because, objectively, it will not involve any state intrusion into privacy. People in the community do not expect to be free from the

---

251 Andrew Butler and Petra Butler *New Zealand Bill of Rights Act: A Commentary* (2nd ed, LexisNexis, Wellington, 2015) at [18.8].

252 *Hamed v R* [2011] NZSC 101, [2012] 2 NZLR 303 at [167] per Blanchard J. Note however, Tipping J's alternative view at [220] that the word "search" should not be defined by reference to expectations of privacy.

> observation of others, including law enforcement officers, in open public spaces such as a roadway or other community-owned land like a park, nor would any such expectation be objectively reasonable. The position may not be the same, however, if the video surveillance of the public space involves the use of equipment which captures images not able to be seen by the naked eye, such as the use of infra-red imaging.

While the last words of the above quote might change the application of s 21 (i.e., on the basis that AFR captures images "not able to be seen by the naked eye"), it is not clear that this is actually the case. Generally AFR captures the same images as the eyes – the difference is that it can use this in new ways (e.g., to very quickly recognise and associate an individual with existing information).

The assumption that things able to viewed normally in public spaces will not give rise to a reasonable expectation of privacy should be challenged.[253] As Elias CJ underlined in her part-dissenting judgment in *Hamed*, this position creates a rigid and unhelpful test for protecting against state intrusion into private affairs which fails to account for technological developments and fluid expectations of privacy among the public.[254] Beswick likewise argues that "the accuracy and targeted nature of surveillance devices, and the permanence of recordings, rarely make them comparable to human perception and memory" and that the Court's approach suggests those in public spaces are unprotected "regardless of the invasiveness of any surveillance devices employed by the state".[255]

However, the presumption that privacy interests should be assessed according to a public/private dichotomy is also reflected in the Search and Surveillance Act 2012 ("**SSA**"). In particular, the SSA only requires enforcement officers to obtain a warrant for visual surveillance where there is surveillance of private activity (i.e., where the persons concerned would expect it to remain private) or continued surveillance in and around a person's home.[256] This suggests that PSAs that deploy AFR in public places may capture

---

253 See Samuel Beswick "Perlustration in the Pathless Woods: *Hamed v R*" (2011) 17 Auck U L Rev 291 at 297 - 299. Commentators have also challenged this public/private divide under the "reasonable expectation of privacy" test applied in tort law: see N A Moreham "Privacy in Public Places" (2006) 65 Cambridge Law Journal 606; and A Jay McClurg "Bringing Privacy Law out of the Closet: A Tort Theory of Liability for Intrusions in Public Places" (1995) 73 North Carolina L Rev 989. See, on the other hand, the Court's view in *Hosking v Runting* [2005] 1 NZLR 1 (CA).

254 *Hamed v R*, above n 252, at [12].

255 Beswick, above n 253, at 297 - 298.

256 Search and Surveillance Act 2012, s 46.

and use individuals' faces for various purposes with little restraint – whether or not the use case relates to the detection or prosecution of crimes (i.e. enforcement purposes), or something more mundane.

On the other hand, the right to be free from unreasonable search and seizure could apply where algorithms are used to gather social media or other data accessible on the internet via an intermediary, and this collection is for enforcement purposes. For example, the right might apply in circumstances where a provider has relationships with social media companies to access user data, and creates inferences about these users which are passed on to a PSA for enforcement purposes (imagine, say, if ACC took prosecutions for fraud partly on the basis of social media data suggesting that clients were not affected by an injury). In these cases, an affected person might be able to rely on s 21 if he or she was able to establish that there was a reasonable expectation of privacy in the data that was used.[257] However, where data is only collected from "public" online forums or profiles, the dichotomy above implies this expectation cannot exist.

Do these issues require change? PSA use of surveillance technologies for non-enforcement purposes is best addressed by amending the definition of "collect" under the Privacy Act, rather than by widening the focus of s 21 of NZBORA. However, where s 21 applies, this thesis argues the courts *must* take a more contextual approach to where a reasonable expectation of privacy can arise. Evolving technologies and expectations of privacy suggest that existing protections against unreasonable search and seizure based on a traditional "small data" approach risk allowing the state to intrude to an unreasonable extent through a vast array of big data tools.[258] Moreover, unless there is adequate oversight over the surveillance tools used by enforcement, the much greater degree of data and tools available to officers may mean that the traditional standard of "reasonable suspicion" in reality becomes lower.[259] All this suggests the courts should acknowledge that reasonable expectations of privacy can and frequently will arise in so called "public spaces", and that

---

257 *R v Alsford* [2017] NZSC 42, [2017] 1 NZLR 710.

258 As Joh notes: "The exercise of surveillance discretion in traditional policing attracts little attention from judges or legal scholars. Why? The answer is likely to be because 1) we assume that Police should possess such powers, and 2) even if theoretically worrisome, surveillance discretion is a power greatly limited in practice. After all, police investigations typically only focus on a limited number of persons because of practical limitations imposed by resources and technology. But those assumptions will become outdated when the police possess the tolls to exercise automated surveillance discretion on a massive scale." See Joh, above n 23, at 17.

259 See Ferguson, above n 26, at 334.

thought should be given to whether the current approach under the SSA – mirroring this presumption – remains fit for purpose.

G Informational rights: practical considerations

The above discussion has explored the legal boundaries of informational rights available to someone subject to an algorithmic decision. However, because it highlights the actual efficacy of these legal protections to respond to algorithmic harms, it is worth briefly focusing on a number of practical considerations.

1 Privacy Act

We start with the Privacy Act as the most likely source of redress. First, the Privacy Act is limited to "personal information" – therefore limitations on use and collection have no impact on the great swaths of data that can be sourced anonymously by PSAs and potentially mined for insights. Likewise, the Privacy Act will not be able to respond to data use which is ostensibly about places or resources but which has direct effects on individuals within certain categories (e.g., with place-based policing or other tools designed for the purpose of resource allocation).

Second, remedies under the Privacy Act hinge upon harm to individuals. This means it is unlikely to be effective for disaggregated forms of harm that occur across a group but which may not meet the standard for an interference with privacy. There is no concept of representative actions under the Privacy Act. While the Privacy Commissioner's functions include inquiring into any matter that could lead to the privacy of individuals being infringed – and so suggest some scrutiny of group harms – this function is not accompanied by any power to enforce sanctions or remedies against offending agencies.[260] It is also heavily resourcing-dependent. For this reason, chapter six therefore proposes a regulatory model better suited to avoiding these diffuse harms from algorithms.

Third, it can be a long time before remedies under the Privacy Act are actually accessible, and damages judgments are not usually substantial. In most cases an individual will first have to show that there has been an interference with privacy (i.e., harm), [261] rather than

---

[260] Section 13.

[261] Section 66.

simply a breach of an IPP.[262] The individual will then have to participate in a process whereby the Privacy Commissioner investigates the complaint and acts as "conciliator".[263] This may lead to a financial or other settlement (e.g., an apology), and an assurance as to the agency's ongoing conduct – but not always. Where there is a settlement, the Commissioner may refer the matter to the independent Director of Human Rights Proceedings ("**Director**") to seek a range of civil remedies (including damages) in the HRRT on behalf of the complainant,[264] or complainants may independently take proceedings (this time at their own cost).

This process can be longwinded. While the Privacy Commissioner does a very good job of disposing of most privacy complaints, the HRRT often provides delayed and meagre justice. As of November 2017, the HRRT's chair noted that cases typically take between 22 to 28 months from the date of filing to the first hearing, and described this delay as "unconscionable".[265]

Moreover, in many cases the current process provides only small monetary awards which are unlikely on their own to truly compensate an affected individual or to provide a strong incentive for compliance with the Privacy Act. Most awards for intrusion into privacy will sit around $2,000 to $20,000[266] – an amount that will almost always be less than the cost of litigation (where the Director does not support the claim), and in an environment where costs judgments are often meagre.[267] Although the Act caps damages at the same levels as the District Court (currently $350,000),[268] the highest damages award to date is

---

262 Section 11.

263 Sections 69 - 80.

264 Part 8.

265 Letter from Rodger Haines QC (Chairperson of the Human Rights Review Tribunal) to Andrew Little (Minister of Justice) regarding the Human Rights Review Tribunal (3 November 2017).

266 Although somewhat dated now, see Katrine Evans "Show Me the Money: Remedies Under the Privacy Act" (2005) 36 VUWLR 475, which provides a helpful breakdown of the damages awarded by the HRRT up until the mid 2000s. More recent cases bear this out: see *Director of Human Rights Proceedings v Crampton* [2015] NZHRRT 35 and *Tapiki and Eru v New Zealand Parole Board* [2019] NZHRRT 5 where there were awards of $18,000, $16,000 and $12,000. In *Director of Human Rights Proceedings v Slater* [2019] NZHRRT 13, the plaintiff was awarded only $70,000 in case involving severe humiliation, severe loss of digntity, and severe injury to feelings. See also the discussion of awards in *Hammond v Credit Union Baywide* [2015] NZHRRT 5 from [168] onwards.

267 For example, in *Hammond* the complainant was not entitled to costs as a lay litigant. See *Hammond Credit Union Baywide*, above n 266, at [190].

268 Subject to exceptional cases. See Privacy Act 1993, s 89; Human Rights Act 1993, s 92Q; and District

$168,070.88.[269] Moreover, while the Privacy Act does allow the HRRT to order the defendant to do or stop doing an infringing act and to take steps to redress loss,[270] in many cases the privacy interference will have come and gone long ago – making the adequacy of damages remedies highly important.

2 Unreasonable search and seizure

It is also useful to consider the practical limitations of pursuing a claim of unreasonable search or seizure.

First, in cases where reasonable expectations of privacy arising under s 21 of NZBORA have been breached, the affected party is only likely to have limited remedies. The presumptive remedy is to have evidence of any unlawful search excluded in any prosecution. However, this turns on the application of s 30 of the Evidence Act 2006, which requires a balancing test based on whether "exclusion of the evidence is proportionate to the impropriety". Unless the intrusion is egregious, in many cases the evidence will remain admissible.

Moreover, damages are an unlikely remedy. While NZBORA *Baigent* damages have been available in rare cases for a breach of s 21[271] – indeed *Baigent's Case* itself concerned an unreasonable search and seizure[272] – the courts have confirmed that exclusion of evidence will be the usual option.[273] Moreover, in those cases where *Baigent* damages have been available, the scope now appears to be limited to the low tens of thousands of dollars – what one prominent lawyer has called "so small as to be derisory" and devaluing of human rights.[274]

Courts Act 2016, ss 74 - 79.

269 *Hammond Credit Union Baywide*, above n 266.

270 Section 85.

271 See *Wilson v NZ Customs Service* (1999) 5 HRNZ 134 (HC); *Small v Attorney-General* (2000) 6 HRNZ 218 (HC); and *Forrest v Attorney-General* [2012] NZCA 125, [2012] NZAR 798.

272 *Simpson v Attorney General* [1994] 3 NZLR 667 (CA) [*Baigent's Case*].

273 *Attorney-General v Van Essen* [2015] NZCA 22.

274 Dr Rodney Harrison QC "Remedies for Breach of the New Zealand Bill of Rights Act 1990: The New Zealand Experience – Recognising Rights While Withholding Meaningful Remedies" (paper presented to the New Zealand Law "Society Using Human Rights Law in Litigation" Intensive Conference, June 2014) 107 at 116.

Therefore, while a s 21 case does not require a party to step through the same process described for the Privacy Act – a party will typically go straight to the District Court or High Court – the cost of this process may be significant, and the redress low.

H Informational rights: conclusion

The foregoing discussion shows that informational rights can provide limited remedies for algorithmic decisions. Rights of access to information may allow persons to understand how algorithms work and can affect decisions about them, and can provide an incentive for agencies to comply with best practice. Likewise, the requirement for agencies to take steps to ensure that information is accurate is likely to provide the most relevant and significant remedial avenue for someone who has suffered an interference with privacy. Meanwhile, s 21 of NZBORA can respond to uses of algorithms by enforcement agencies that intrude upon reasonable expectations of privacy.

However, a number of issues unnecessarily hinder individuals' prospect of accessing remedies, and changes should be made to ensure that these laws do in fact achieve the underlying policy goals on which they are based. First, simple tweaks to the Privacy Act can ensure that algorithmic outputs can be personal information and that "collection" includes passive collection. This would enhance protection under IPP 8 by requiring reasonable steps to ensure *outputs* are accurate, and would ensure that unfair surveillance activities can be addressed by IPP 4. The OIA's objectives of transparency in public decision-making,[275] as well as the public's comfort with algorithmic decisions, would be significantly enhanced by new rights for individuals to receive notice of an algorithmic decision and to access human review. The shape of an existing "right to reasons" should also be clarified as it applies to algorithms, to ensure a proportional response is available. Lastly, changes should relax the public/private dichotomy for reasonable expectations of privacy under s 21 of NZBORA in appropriate situations, and parallel changes should be made to the SSA to reflect this.

Even with these improvements, however, the informational rights above only provide a limited mechanism to address algorithmic harms. From a practical perspective, the Privacy Act's monetary remedies are unlikely to be significant and the process can be longwinded. Likewise, remedies under s 21 of NZBORA will rarely involve damages and exclusion of

---

275 OIA, s 4; LGOIMA, s 4.

evidence is not guaranteed even if a reasonable expectation of privacy has been breached. The OIA provides rights to access information, but is not remedial in nature.

Importantly, these remedies respond only to *individual-focused* harms and not to low level group harms. The right to sue for an interference with privacy under the Privacy Act requires the affected individual to meet a material standard of harm.[276] While this is appropriate to avoid vexatious claims, it also illustrates how the rights-centred nature of remedies under the Privacy Act fails to respond to large-scale but otherwise minor privacy harms which can still have a broad impact upon groups in society (e.g., if individuals are systematically mischaracterised). Similarly, rights under s 21 of NZBORA will not typically respond to low level harms that erode the broader privacy commons, but which are not a significant intrusion by an enforcement agency into the privacy of one individual alone.

Bearing this in mind, chapter six discusses how a new regulatory framework is able to address these issues. However, for now we turn to discussion of the HRA.

---

276 Section 66.

IV Chapter four: Rights against Discrimination under the Human Rights Act 1993

A Overview

This chapter investigates the ability of the HRA to provide recourse where an algorithmic decision discriminates in a way that causes harm to individuals.

As we will explore, claimants will face a range of hurdles – some specific to algorithmic decisions, some not – in establishing a claim under the HRA. Perhaps most significant will be the grey line dividing decisions where there is substantial reliance on an algorithm (a "**directed decision**") and those where a decision-maker relies on a range of factors, just one of which is the algorithm's output (an "**informed decision**"). Other issues for a claimant concern the "indirect" nature of most discrimination claims, the challenges of showing material harm to a group, and practical barriers to taking a HRA claim. However, this chapter also suggests that where causation and harm are sufficient for a claim to establish a prima facie discrimination case, the PSA will often face an uphill task in defending the claim by asserting that the discrimination should be "justified" in a free and democratic society – particularly where there are no records anticipating how an algorithm might cause this discrimination.

B Scope of HRA

The HRA provides remedies where a person has suffered discrimination on the basis of a prohibited ground of discrimination ("**prohibited ground**"). The prohibited grounds are found in section 21 of the HRA and cover discrimination on the basis of sex, marital status, religious belief, ethical belief, colour, race, ethnic or national origin, disability, age, political opinion, employment status, family status and sexual orientation.[277] Section 19 of NZBORA constitutionally embeds the obligation for the state to give equal protection and equal benefit of the law by providing that "everyone has the right to freedom from discrimination" on the prohibited grounds.

Most discrimination claims based on a PSA's actions must be taken under Part 1A of the

---

277 Human Rights Act 1993 [HRA], s 21.

HRA.[278] Part 1A applies to acts or omissions of persons or bodies referred to in section 3 of NZBORA, which includes any entity that exercises a "public function, power or duty" under law.[279] Remedies will be available if a PSA's algorithmic decision is inconsistent with s 19 of NZBORA (i.e., protection against discrimination on a prohibited ground) because it:[280]

> (a) limits the right to freedom from discrimination affirmed by that section; and
>
> (b) is not, under section 5 of [NZBORA], a justified limitation on that right.

To prove that an algorithmic decision has caused unjustified discrimination, the complainant will need to show that:[281]

> (1) there is differential treatment or effects as between persons or groups in analogous or comparable situations on the basis of a prohibited ground;
>
> (2) the differential treatment has a discriminatory impact, that is, when "viewed in context, it imposes a material disadvantage on the person or group differentiated against"; and
>
> (3) in accordance with section 5 of NZBORA, the differential treatment is not a limitation on the right to be free from discrimination found in s 19 of NZBORA that "can be demonstrably justified in a free and democratic society".

For the first two questions, the onus will lie on the plaintiff. However, once the plaintiff has made out a prima facie case that discrimination exists on the basis of a prohibited ground, and this imposes a material disadvantage, it is for the defendant PSA to prove that the measure is a justified limitation on the right.[282] A decision is typically only justified if

---

278 Some claims can, however, be taken under Part 2 to the extent they relate to discrimination in employment matters, racial disharmony and social and racial harassment, or victimisation. See HRA, s 21A.

279 Section 20J.

280 Section 20L.

281 See *Ministry of Health v Atkinson* [2012] NZCA 184, [2012] 3 NZLR 456 at [55], [109] and [143]; followed in *Ngaronoa v Attorney-General; Taylor v Attorney-General* [2017] NZCA 351, [2017] 3 NZLR 643.

282 See HRA, s 92F; and *Ministry of Health v Atkinson*, above n 281, at [163].

it can pass the standard *R v Hansen* test. That asks:[283]

> (a) Does the limiting measure serve a purpose sufficiently important to justify the curtailment of the right?
>
> (b)
>
> (i) Is the limiting measure rationally connected with its purpose?
> (ii) Does the limiting measure impair the right or freedom no more than is reasonably necessary for sufficient achievement of the purpose?
> (iii) Is the limit in due proportion to the importance of the objective?

C Key challenges: prima facie case

Having sketched out the basic legal framework, we can now examine the issues that could affect a HRA claim based on the use of an algorithm. While the HRA in principle responds to some of our previously discussed examples – such as the racially-biased sentencing scenario – a complainant is likely to face a number of challenges making a prima facie case in relation to an algorithmic decision. These include bringing an indirect discrimination claim, demonstrating material harm and causation, and choosing the correct comparator group.

1 "Indirect discrimination" precedent

Most claims in relation to an algorithmic decision will typically be an "indirect discrimination" claim. Indirect discrimination claims differ from other discrimination claims by focusing on the *effects* of a decision, policy or law, rather than whether that limitation is on its face discriminatory.[284] This kind of claim is likely based on the cases discussed in chapter two: it seems rare that a PSA will use an algorithm for a decision with the knowledge or intention that it gives rise to unfair differential treatment. Rather, it is more likely discriminatory treatment will arise because the algorithm provides outputs which – for unintentional reasons relating to the algorithm's model or data sources – discriminate against a particular group on a prohibited ground.

A first challenge for a complainant in an algorithmic indirect discrimination case will be a

---

283 *R v Hansen* [2007] NZSC 7, [2007] 3 NZLR 1 at [104].

284 *Ngaronoa v Attorney-General*, above n 281, at [119]; and Butler and Butler, above n 251, at [17.12.1].

lack of available precedent to bolster the case. While it is clear that indirect discrimination claims are available under the HRA,[285] none has as yet been successful in New Zealand.[286] The New Zealand indirect discrimination case which has come closest to success is arguably *Ngaronoa v Attorney-General*.[287] While that case is likely to provide some assistance, it demonstrates the high bar to success. There the plaintiffs claimed that a statutory prohibition on prisoner voting discriminated on the grounds of race because of the disproportionately high number of prisoners who are Māori and the resulting disproportionate effect on Māori. The High Court held that no discrimination arose between Māori and non-Māori prisoners because they were equally disadvantaged by the voting ban.[288] However, if instead the Māori voting community was compared with the non-Māori voting community, the Court of Appeal held that there was an indirect difference in treatment (because of the greater proportion of Māori voters generally in prison).[289] Nevertheless, this did not impose a material disadvantage, "affecting as it [did] less than one percent" of Māori as a group.[290]

Moreover, where indirect discrimination claims have been successful in comparable overseas jurisdictions, they have been more straightforward than is likely for an algorithmic decision. For example, in *Eldridge v British Columbia (Attorney-General)*[291] the Canadian Supreme Court found that British Columbia's failure to fund translation services for deaf patients meant that the patients were effectively deprived of access to core medical services due to their disability. Likewise, in *British Columbia (Public Service Employee Relations Commission) v BCGEU*[292] the standard for fitness tests for forest firefighters was discriminatory because it would exclude most women and there was not a clear case this

---

285 Section 65 of the HRA explicitly makes provision for indirect discrimination claims, and its application has been confirmed in recent cases. See *Ngaronoa v Attorney-General*, above n 281, at [111]; *Smith v Air New Zealand Ltd* [2011] NZCA 20, [2011] 2 NZLR 171 at [15]; and *NRHA v Human Rights Commission* [1998] 2 NZLR 218 (HC) at 236.

286 Rather, well-known successful cases have related to matters such as access to funding for parents supporting their disabled children, and resourcing cuts for services aimed at intellectually disabled persons over 65. See *Ministry of Health v Atkinson*, above n 281; and *Attorney-General v IDEA Services Ltd (In Statutory Management)* [2012] NZHC 3229, [2013] 2 NZLR 512.

287 *Ngaronoa v Attorney-General,* above n 281.

288 *Taylor v Attorney-General* [2016] NZHC 355, [2016] 3 NZLR 111; *Ngaronoa v Attorney*-General, above n 281, at [133].

289 *Ngaronoa v Attorney-General,* above n 281, at [147].

290 At [148].

291 *Eldridge v British Columbia (Attorney-General)* [1997] 3 SCR 624.

292 *British Columbia (Public Service Employee Relations Commission) v BCGEU* [1999] 3 SCR 3.

standard was necessary for either men or women. In these situations, the policy was clear and one could draw a direct arc between its application and its discriminatory effects based on a prohibited ground.[293] However, as discussed below, this connection could be harder to draw for algorithmic decisions.

2 Materiality and causation

If a plaintiff can show that an algorithm's outputs correspond with a prohibited ground, they must then prove that a relevant criterion is a "material ingredient" or "operative factor"[294] in the differential treatment. The way algorithms influence decision-making suggests a layered test which is potentially challenging for a plaintiff.

At the first layer, the plaintiff will need to show that differential treatment on a prohibited ground arises from any relevant aspects of the algorithm (e.g., how it is constructed or the data used). From an evidential perspective, this could be challenging to the extent the "reasoning" behind an algorithm is not explainable, because it is a ML algorithm (as discussed above). Nevertheless, this aspect is not insurmountable if it is clear that the *effect* of the algorithm's outputs overall creates differential treatment between groups.

The second layer is arguably more challenging. It asks: has the discriminatory output tainted the overall decision to such an extent that that the decision imposes differential treatment on a group? In this calculation, it will not necessarily matter that there are other criterion taken into account which support the same ultimate result.[295]

Rather, the most important factor is likely to be whether the decision is an algorithmically directed decision (where the algorithm's output has been entirely or substantially relied upon by the decision-maker) or an algorithmically informed decision (where the algorithm's output is one of a number of factors relied upon by the decision-maker). If a decision-maker unquestioningly implements an algorithmic decision or recommendation imposing different treatment based on a prohibited ground, differential treatment will be easy to establish. But, if the algorithmic output is one of many criteria relied upon by the

---

293 For an English example of indirect discrimination, see *Patmalniece v Secretary of State for Work and Pensions* [2011] UKSC 11, [2011] 1 WLR.

294 *CPAG v Attorney-General* [2013] NZCA 420, [2013] 3 NZLR 729 at [53] - [64].

295 The Court of Appeal has noted "the existence of another criterion which may render the person ineligible for assistance does not of itself mean there may not be discrimination on a prohibited ground". See *CPAG v Attorney-General*, above n 294, at [64].

decision-maker, there is again the issue of whether the output was a material ingredient (not to mention the evidential challenge of trying to get the decision-maker to acknowledge this).

A case in point is the Wisconsin Supreme Court decision of *Loomis*.[296] There the Court acknowledged the way in which offender risk prediction software might discriminate on the basis of race. However, the appellant could not prove this because the tool was protected from disclosure by intellectual property protections. Despite this failure of transparency, the Court considered the effect of the tool's risk assessment would not have been "determinative" because it was one of many factors taken into account by the judge.[297] The Court's position appears too broad brush – just because a particular factor is one of many, does not mean it did not materially influence the outcome. It is also at odds with the literature on automation complacency, which suggests judges could be susceptible to undue reliance on algorithmic outputs. However, it does highlight that in the New Zealand context the extent to which a decision-maker nominally retains discretion will be highly relevant, as will the plaintiff's ability to bring convincing expert evidence that the decision-maker was materially influenced by the algorithm and did not act independently for reasons irrelevant to any prohibited ground.

Lastly, if a plaintiff can prove differential treatment, they will also need to show that this imposes a "material disadvantage" upon those subject to that treatment. Again, this question will require clear evidence of the effect on the plaintiff (and those in the applicable group), which may be challenging to collect. Moreover, whether there is material disadvantage (or differential treatment at all) will be affected by the court's choice of comparator.

3 The comparator group

The court's choice of the comparator group is an issue in any HRA claim, and will also be relevant to one concerning a PSA's use of an algorithm. As part of the first step in the legal test, the claimant will need to show that it received a different outcome from another person

---

296 *State v Loomis* 881 NW 2d 749 (Wis 2016). For context, see Katherine Freeman "Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v Loomis" (2016) 18 NCJL & Tech 75; and Frank Pasquale "Secret Algorithms Threaten the Rule of Law" *MIT Technology Review* (1 June 2017).

297 *State v Loomis*, above n 296, at [102] - [110].

or group in similar circumstances.[298] However, the decision about which characteristics define the comparator is often a battleground that can turn a case one way or another.[299] As Elias J has observed, "the choice of comparator is often critical" and can pre-determine the outcome.[300] Rather than operating as a neutral framework for identifying differential treatment, Emmanuel identifies how finding the comparator can be a vehicle for background norms "concerning which persons are to be seen as equal".[301]

While the risk of the "incorrect" comparator is not specific to algorithmic decisions, it is an important consideration. Consider the place-based policing scenario. At first glance, this scenario suggests harms to which the HRA might respond. But what would the comparator be? A plaintiff is unlikely to succeed by arguing that Police treated Pacific Island offenders differently from other ethnicities. Rather, a court is likely to find that Police would also arrest offenders of other ethnicities in the same way for the same offences. However, rather than comparing offenders, it may be better to compare how Pacific Islanders in the neighbourhood are generally treated, compared with those in pre-dominantly Pākehā neighbourhoods. With this different comparator, a plaintiff may actually be successful.

The comparator will naturally affect the question of material disadvantage. As *Ngaronoa* demonstrates, a shift to a broader group might help a party establish differential treatment on the basis of a prohibited ground. But, there is no guarantee that this will help establish material disadvantage. In fact, a broader group may well dilute the ability to bring the evidence necessary to show a high degree of impact. Even with a "better" comparator, the hypothetical scenario above would likely fall short of meeting the test under the HRA. This suggests that the HRA is unlikely to respond to low-level systemic harms that arise from the use of algorithms, even when these discriminate on the basis of a prohibited ground.

D Key challenges: justification analysis

Now we turn to the issues that will be relevant if a plaintiff is able to prove material disadvantage. At this stage, the onus shifts and the PSA will need to prove that this treatment is justified under s 5 of NZBORA. Three factors will be particularly relevant: the

---

298 *Quilter v Attorney-General* [1998] 1 NZLR 523 (CA) at 573.

299 See Asher Gabriel Emanuel "To Whom Will Ye Liken Me, and Make Me Equal? Reformulating the Role of the Comparator in the Identification of Discrimination" (2014) 45 VUWLR 1.

300 *Air New Zealand Ltd v McAlister* [2009] NZSC 78, [2010] 1 NZLR 153 at [34].

301 Emanuel, above n 299, at 5.

extent to which the PSA can establish that the treatment was "prescribed by law"; the PSA's ability to safely navigate the *Hansen* test; and extent to which the court will be willing to provide deference to the PSA given the nature of the decision.

1 Prescribed by law

Section 5 of NZBORA states:

> … the rights and freedoms contained in this Bill of Rights may be subject only to such reasonable limits *prescribed by law* as can be demonstrably justified in a free and democratic society.
>
> (emphasis added)

The following discussion shows how, where an algorithm has unintentionally produced differential treatment between groups, it will be particularly hard for a PSA to assert that this differential treatment was a limitation on s 19 of NZBORA "prescribed by law". If the PSA is unable to prove this point, its defence will fail – and the claimant will succeed – without any need to step through the *Hansen* test.

In short, to be prescribed by law a limit must "be identifiable and expressed with sufficient precision" and "must be neither ad hoc nor arbitrary and their nature and consequences must be clear, although the consequences need not be foreseeable with absolute certainty".[302] These comments would seem fatal in typical cases where algorithmic discrimination occurs, because this discrimination is usually an unintentional outcome of how a policy *is applied* in relation to specific decisions, rather than an intentional aspect of an accessible policy. For example, if there is a policy to provide more restrictive parole conditions for those identified as having a higher risk of offending, but an algorithmic risk analysis systematically and wrongly discriminates against women because of algorithmic

---

302 *Ministry of Health v Atkinson*, above n 281, at [183]. In *Attorney-General v IDEA Services*, above n 286, at [177] the Court quoted the views of Professor Peter Hogg: "The requirement that any limit on rights be prescribed by law reflects two values that are basic to constitutionalism or the rule of law. First, in order to preclude arbitrary and discriminatory action by government officials, all official action in derogation of rights must be authorised by law. Secondly, citizens must have a reasonable opportunity to know what is prohibited so that they can act accordingly. Both these values are satisfied by a law that fulfils two requirements: (1) the law must be adequately accessible to the public, and (2) the law must be formulated with sufficient precision to enable people to regulate their conduct by it, and to provide guidance to those who apply the law."

bias, it seems hard to say that this discriminatory outcome is prescribed by law. The discrimination is not articulated in the policy itself, but rather how it is put into practice by the algorithm.

However, in *IDEA Services* the High Court suggested that in cases where a statute confers a discretion, it is the "decision (not the statute) which is subject to the test of whether it is reasonable and can be demonstrably justified in a free and democratic society."[303] This would suggest each algorithmic decision can be analysed on its own merits as to whether it is "prescribed by law" and if so, whether it meets the *Hansen* test. But, a decision still needs to have appropriate precision and accessibility, although these are not "required in absolute terms … and the particular context will be relevant to the degree to which precision and accessibility are required."[304]

Because *IDEA Services* dealt with a decision to establish a consistent position, there is a good argument that it should be distinguished from the kind of case-by-case decisions made when an algorithm is used to help assess an individual. The cases cited in support of *IDEA Services* also typically concern these kind of "one-off" decisions.[305]

But even if this view is not accepted, it will be hard for a PSA to prove a measure was prescribed by law unless an explanation of how the algorithm works is available. Where an algorithmic output is relied upon and causes the person material disadvantage based on a prohibited ground, the law will not be sufficiently precise and accessible if a person cannot access how the algorithm has created that output. This is particular risk for ML algorithms, given their opacity. This is likely to be the case *even if the algorithm discriminates in a consistent way*.

The upshot of the above is that a defensible argument that a limitation is prescribed by law is probably only available to a PSA if it publishes a clear and accessible explanation of how the algorithm produces its outputs. This may be sufficient where the algorithm acts relatively consistently and in accordance with that explanation. However, the requirements of precision and accessibility may be extremely hard to satisfy in the case of ML

---

303 At [186].

304 At [190].

305 These are: *Wynberg v Ontario* (2006) 269 DLR (4th) 435 (ONCA); *Christchurch International Airport Ltd v Christchurch City Council* [1997] 1 NZLR 573 (HC); *Federated Farmers of New Zealand Inc v New Zealand Post Ltd* [1992] 3 NZBORR 339 (HC); and *Slaight Communications Inc v Davidson* [1989] 1 SCR 1038.

algorithms, where the criteria and weightings for a given outcome can move constantly, and because the algorithm's technical language is directly unreadable. In these cases, if the algorithm becomes too inconsistent with any explanation, the plaintiff will have a strong argument that the PSA has no grounds for a justification defence.

2 Hansen *test*

But, assuming that an algorithmic decision is "prescribed by law", the question of justification will again be affected by whether the case concerns an algorithmically directed decision (i.e., one substantially relied up) or simply an informed decision. As outlined above, the *Hansen* test requires the PSA to demonstrate that relying on the algorithm's (discriminatory) output would: serve a sufficiently important purpose; impose a limit rationally connected with the purpose; impair the right no more than reasonably necessary; and be proportional.

Algorithmically directed decisions seem highly unlikely to meet the *Hansen* test unscathed. A key aim of s 5 of NZBORA is to embed a "culture of justification".[306] This hints at the need for decisions to be made with due particularity and with necessary consideration of the impact on the rights of citizens to ensure they can be justified. This is also reflected in the Crown having the burden of proving that a limiting measure meets the s 5 test.[307] This rationale suggests any automated culture of justification would be a misnomer – because even a highly accurate algorithm is unlikely to be able to get the full terrain and contextual understanding of how its decisions will affect individuals.

Moreover, in algorithmically directed cases it seems unlikely that a PSA will be able to show that any unintentional discrimination serves a wider, sufficiently important purpose and that this discrimination is rationally connected to it. At best, the PSA would need to have determined that whatever discrimination could arise was an unavoidable cost of meeting sufficiently important policy goals via automation – *without knowing what this*

---

306 As Butler and Butler note: "A 'culture of justification' means a culture in which citizens are entitled to call upon the provision of reasons for measures that affect their rights, are entitled to challenge those reasons, and in a sense more importantly, are entitled to expect that in advance of impairment thought will have been given to the reasonableness of a particular limit. The culture of justification contributes to principles of good government, such as transparency, accountability, rational public policy development, attention to differing interests, and so on." See Butler and Butler, above n 251, at [6.8].

307 At [6.8.2]

would look like. Effectively, the PSA would be trying to justify recklessness as to any unintentional harm. And even if the PSA could show any discrimination helped to achieve a sufficiently important purpose that is rationally connected, how would the PSA show the discrimination "impairs the right no more than reasonably necessary" and is proportional? Would it be enough for the PSA to show it had carefully vetted the algorithm and the input data, and had taken steps to ensure it would generally produce the "right" results? Arguably this would still fall short of the particularity required of s 5 and also serve to diffuse responsibility for individual decisions by pointing to a general process of algorithmic hygiene.

However, plaintiffs will have a harder time where there is an algorithmically informed decision. Here, the decision-maker should be in a stronger position to bring evidence indicating why the decision accords with the *Hansen* criteria, and to explain the extent to which the algorithm contributed to the outcome. To the extent that the decision-maker can also articulate his or her understanding of how the algorithm produces its outputs, its accuracy and validity with specific groups, and its understanding of the criteria, and the likely impact of false positives or negatives, this will help the PSA.

3 Deference

Lastly, a court's degree of deference to the PSA, based on the subject matter, will also be relevant to whether any discrimination is justified. In essence, a court is more likely to defer to a PSA under the s 5 test when a limitation involves "major political, social or economic decisions", but will adopt a greater "intensity of review" in matters involving "substantial legal content".[308]

For example, a PSA may receive the benefit of the doubt where the rationale for using algorithms is to cut costs or increase efficiencies in difficult areas of social and fiscal policy. So, where an algorithm suggests MSD should intervene with a family because its risk model suggests a higher chance of child abuse, or an algorithm otherwise helps prioritise how resources should be deployed between competing groups, the courts may be more deferential. We can expect that this question of scrutiny will be particularly relevant to the second and third limbs of the *Hansen* test: in short, whether a rights limiting measure was one of a range of options reasonably available to the PSA to take, and whether the limit

---

308 *R v Hansen*, above n 283, at [116]. See also *CPAG v Attorney-General*, above n 294, at [91].

was proportional. A court may be more willing to entertain a higher degree of harm arising from the use of algorithms if it considers the area one in which it is not institutionally well-equipped to judge.

On the other hand, there are a number of areas more closely within the courts' institutional competence where closer scrutiny may be adopted – for example if the matter concerns the use of algorithms for the purpose of sentencing, gathering evidence for prosecution, or to decline individuals' statutory entitlements (for example, someone's access to healthcare through ACC).

The courts' inclination to intervene is also likely to be tempered by the limits of uncertainty in policy-making (and any decision-making). New policies are often adopted in an "environment of scientific or other uncertainty and complexity" and will entail an element of speculation.[309] Therefore similar allowances should be made for the use of algorithms. Likewise, although the inferential nature of algorithmic decisions creates an obvious tension between the individual and their representative group, this tension still exists generally in public decision-making. The comments of Butler and Butler should be kept in mind:[310]

> …the focus of public policy considerations is usually discussed in terms of a "typical" scenario thought to be representative of that problem. Reponses are devised to respond to that paradigmatic scenario… care needs to be taken in constructing standards of reasonableness: an insistence on exactitude, requiring the legislature/executive to avoid any underinclusiveness or overinclusiveness may be premised on unreal expectations of what the regulatory system can be expected to deliver.

E HRA: practical considerations

Last, to better assess the HRA's ability to respond to algorithmic harms in practice, we should quickly examine the practical limitations of a bringing a HRA claim. The below shows that a HRA claim involves practically the same process as outlined above in relation to the Privacy Act, and many of the same obstacles.

---

309 Butler and Butler, above n 251, at [6.9.13]. For example, the authors of an ethical review into MSD's proposal to use algorithms to infer children at risk of child abuse admitted that whether the benefits outweighed the risks were unknowable until a trial was conducted. See Blank et al, above n 197, at 2.

310 At [6.9.17] - [6.9.18].

What are these considerations? First, before seeking a judicial remedy,[311] a complainant must complain to the Human Rights Commission ("**HRC**"). Like the Privacy Commissioner, the HRC will endeavour to facilitate a settlement between the parties,[312] and where this is not reached the Director may take the matter to the HRRT on behalf of the complainant.[313] The complainant may also independently take the matter in the HRRT.[314]

Second, the process for claims is long-winded. HRA cases are heard at first instance in the HRRT, like privacy cases, and the comments from chapter three apply equally.

Thirdly, the remedies available will depend on the nature of the impugned act or omission. Where the complaint relates to an enactment, the only remedy available is a declaration that the enactment is "inconsistent with the right to freedom from discrimination affirmed by section 19" of NZBORA.[315] However, it seems unlikely a person affected by an algorithmic decision will be complaining about the requirements of the enactment, but rather the decisions(s) of a PSA under a discretion or policy. In this case, most of the remedies available are the same as under the Privacy Act and stretch from damages to orders to do or stop doing an action in relation to discriminatory treatment.

Fourth, the complexity and level of evidence required potentially imposes a high barrier to claimants unless the Director supports proceedings. As should be clear from the tests outlined above, a complainant will need to have evidence demonstrating a relatively high standard of causation and harm to individuals in the relevant group. To contest the PSA's justification position, a complainant will also typically need to comb through reams of evidence from the PSA relating to the rationale for the use of the algorithm in decision-making, and why it is a rational, proportional and reasonably limited measure. The complexity and extent of evidence in cases like *Atkinson*, *IDEA Services* and *CPAG* is instructive.[316] All of this work takes time and money.

Fifth, HRA claims may, however, potentially lead to higher damages awards than under

---

311 Civil proceedings under Part 1A may be commenced under s 92B of the HRA.

312 HRA, ss 77 and 83.

313 HRA, s 84.

314 HRA, s 92B.

315 HRA, ss 92I - 92K.

316 *Ministry of Health v Atkinson*, above n 281; *Attorney-General v IDEA Services*, above n 286; and *CPAG v Attorney-General* above n 294.

the Privacy Act. For example, in *Spencer* the High Court upheld a pecuniary loss award of $233,091 for a single claimant – more than has ever been given under the Privacy Act.[317] The availability of significant damages is also supported by the purpose of Part 1A to provide "effective remedies".[318] As Barnett outlines, the guiding principle underlying the Part 1A damages regime is that of *resitutio in integrum*: that compensation should put an affected party back to its former position.[319] This accords with the words of one of the Bill's architects, the Hon Margaret Wilson, that Part 1A "provides individuals with protection from Governments exercising arbitrary or discriminatory power" and "heralds a new era of public sector accountability".[320] Therefore, for cases which can demonstrate real harm, there is good reason to think substantial damages may be available.

Lastly, the HRA has additional jurisdiction for the court to order the PSA to implement new programmes or policies to help compliance with the HRA.[321] This could act as an important mechanism to correct and improve the PSA's use of algorithms in decision-making.

F HRA: conclusion

This chapter has shown that the HRA can provide potentially significant remedies for those who are affected by algorithmic decisions. Like the Privacy Act, the HRA on its face appears to address some of the potential harms typically associated with decision-making algorithms – such as systematically inaccurate or biased outputs that disadvantage those in a particular group. The HRA *will* respond where it is easy to show material disadvantage arising from different treatment, and this treatment arises substantially from the use of an algorithm. This is particularly likely to be the case if the discrimination is unintentional and/or the discriminatory decision is solely automated.

However, a number of factors impact the efficacy of the HRA in responding to algorithmic harms. First, because algorithms will rarely be used with an express intention to treat one

---

317 *Spencer v Ministry of Health* [2016] NZHC 1650, [2016] 3 NZLR 513; and *Spencer v Ministry of Health* [2017] NZHC 291.

318 Human Rights Amendment Bill 2001 (152-1) (explanatory note) at 1.

319 Peter Barnett "Remedies for Discrimination by Government under Part 1A of the Human Rights Act 1990" (paper presented to the New Zealand Law "Society Using Human Rights Law in Litigation" Intensive Conference, June 2014) 135 at 136 - 137.

320 (11 December 2001) 597 NZPD 13759, cited in Barnett, above n 319, at 137.

321 Section 92I.

or more groups differently on the basis of a prohibited ground, almost all algorithmic discrimination cases are likely to be based on an indirect discrimination claim. While these claims are recognised by the HRA, and have been successful overseas, none has been successful in New Zealand and a claimant will have to bring a novel case with little precedent.

Second, the likelihood of success will depend greatly on how the algorithmic tool is used. If decisions are largely delegated to or automated via the tool (a directed decision), it will be easier to show that a discriminatory algorithmic output was a material contributor to a decision, and potentially also that the exercise of public power was not "prescribed by law" (and so, cannot be a "justified" infringement on the right to be free from discrimination). On the other hand, if an algorithm merely informs the decision-maker as one of a number of factors taken into account (an informed decision), it will be harder for the plaintiff to show the algorithm helped cause the discriminatory outcome.

Moreover, courts will have regard to the context in which algorithms are used. Where there are important consequences for the individual (e.g., a longer term of imprisonment, or an order for deportation), we can expect that courts will be less willing to accept the risks of harmful false positives or negatives as justifiable limits on the right. However, in other areas this may be different, given the efficiencies and public-policy benefits that can arise from algorithms.

Remedies under the HRA also suffer from similar drawbacks to those described for the Privacy Act. In particular, the HRA is unlikely to provide a suitable mechanism for relatively diffuse harms because, even if differential treatment can be established, it may not chin the "material disadvantage" threshold. For example, *Ngaronoa* suggest the HRA is unlikely to respond in the place-based policing scenario or "resource" focused claims which impact communities.

Moreover, the process to access remedies can be long and difficult. Although monetary damages under the HRA may be more significant than under the Privacy Act, the cost and barriers of bringing a claim mean it is only like to be attractive in serious cases.

However, this thesis does not suggest changes are needed to the HRA to make it respond to algorithmic harms. Concerns about diffuse harms are best addressed by the regulatory response suggested below. The need to show causation will still often be a barrier for claimants in informed decision cases – but ultimately this is likely to be resolved through the use of expert evidence that is able to demonstrate the impact of automation bias on

decision-makers. Likewise, although a claimant will face a challenge bringing a novel indirect discrimination claim based on the use of an algorithm, meritorious claims are still likely to be backed by the resources of the Director.[322] And lastly, the deference shown by the court is ultimately a reflection of the institutional relationship between the courts and the other branches of government – one that is appropriate, even if not always helpful for a claimant. Instead of changing the HRA, any remaining gaps in the matrix of legal accountability for these harms should be addressed by a new regulatory model, as outlined in chapter six.

Now, we move to the last chapter on remedial avenues for algorithmic harm: judicial review.

322 HRA, s 90.

V Chapter Five: Judicial Review

A Overview

This chapter looks closely at whether judicial review can respond to algorithmic harms and, if so, any changes that should be considered to improve its applicability to PSAs' use of algorithms.

As we will explore, judicial review provides potentially the most viable and effective path for a person to challenge the use of algorithms by a PSA. A range of overlapping grounds of review may be available, given the way algorithms can lead to "autopilot" decision-making and reduce individuals' ability to be heard by the decision-maker. Moreover, unlike the HRA or Privacy Act, an applicant for judicial review does not need to establish harm to bring a claim. This allows judicial review to address both a particular decision which has a significant impact on the individual, and a PSA's *practice* of algorithmic decision-making which contributes to system-wide but low level harm. However, this chapter also shows how judicial review is not without drawbacks - for example, the cost of review and the non-compensatory and process-driven nature of remedies.

Ultimately though, judicial review is the most far-reaching available remedy and should help ensure accountability for algorithmic decision-making. In fact, in some cases judicial review is possibly *too* responsive to algorithmic use, and risks hindering legitimate use of algorithms for beneficial purposes. Given this issue, chapter six proposes a new regulatory model that legitimises PSA use cases within appropriate constraints, without losing the ultimate benefits of algorithmic decision-making.

For now, however, we turn to the basics of judicial review.

B Judicial review: basic outline

Judicial review stems from the High Court's inherent jurisdiction to supervise the actions of public decision-makers to ensure these are lawful. While there are numerous grounds of judicial review – and the High Court will exercise its powers whenever the circumstances demand it[323] – its scope can be neatly captured by Cooke P's often repeated mantra that

---

[323] Hence, the so-called "innominate ground" arising from Lord Donaldson MR's comments in *R v Panel on Take-overs and Mergers, ex parte Guinness Plc* [1990] 1 QB 146, [1989] 1 All ER 509 (CA).

decisions must be made "in accordance with law, fairly and reasonably".[324] These requirements fall under the general principle that public bodies must comply with the rule of law.[325] Hence a decision will be unlawful if a decision-making power is interpreted over broadly, or if the requirements of procedural fairness implied by the law are not observed, or if the decision is of a nature Parliament would never have intended to allow (and so is substantively unreasonable).[326] An application for judicial review will be made under the Judicial Review Procedure Act 2016.[327]

Judicial review should almost always be available where a PSA makes an algorithmic decision, because all exercises of public powers are in principle reviewable.[328] Algorithmic decisions should easily meet the "public law" focus or element test for review,[329] and the generous approach to standing in New Zealand judicial review cases[330] means that affected persons will be able to bring a case.

However, the "justiciability" of the claim and the court's "intensity of review" are further contextual factors that could limit the availability or impact of judicial review – and, therefore, its utility to regulate improper use of algorithms. These matters are keenly fought territory for administrative law scholars, and so the purpose here is simply to acknowledge their relevance, rather than delve into a detailed analysis of their scope.

First, a claim will need to be "justiciable". A non-justiciable matter is:[331]

> … one in respect of which there is no satisfactory legal yardstick by which the issue can be resolved. That situation will often arise in cases into which it is also constitutionally inappropriate for the Courts to embark.

The algorithmic decisions which are the subject of this thesis should usually meet this test,

---

324 *New Zealand Fishing Industry Association Inc v Minister of Agriculture and Fisheries* [1988] 1 NZLR 544 (CA) at 552.

325 *Tannadyce Investments Ltd v Commissioner of Inland* Revenue [2011] NZSC 158, [2012] 2 NZLR 153 at [3] per Elias CJ and McGrath J.

326 Francis Cooke "Judicial Review" (New Zealand Law Society seminar, May 2012) at 3 - 4.

327 The Act re-enacts Part 1 of the Judicature Amendment Act 1972.

328 *Ririnui v Landcorp Farming* [2016] NZSC 62 at [1] and [89].

329 *Air Nelson Ltd v Minister of Transport* [2008] NZCA 26, [2008] NZAR 139 at [33].

330 *Ririnui v Landcorp Farming*, above n 328, at [91]; *Kim v Prison Manager Mount Eden Correctional Facility* [2012] NZSC 121, [2013] 2 NZLR 589 at [76].

331 *Curtis v Minister of Defence* [2002] 2 NZLR 744 (CA) at [27].

because they will typically involve rights, entitlements and sanctions enforced by the state against individuals. In cases where the use of algorithms are highly abstracted from individuals or the decision is best challenged through the democratic process[332] – such as public finance decisions[333] – that position may be different.

The High Court may also adopt a varying "intensity of review" based on the subject matter at hand.[334] This is related to "deference" under the HRA, and suggests the Court will assume a more "intense" review – or take a "hard look" or adopt a stance of "anxious scrutiny" – where a case concerns human rights matters in which the Court is expert.[335] This approach may apply in related areas such as immigration or deportation decisions, asylum claims, extradition cases, and possible Treaty of Waitangi cases.[336] Michael Taggart famously described a "rainbow of review", spanning these kinds of rights-related cases at one end and more typical "public wrongs" at the other.[337]

Lastly, the overall scheme in which a decision is made will be important. For example, the Court will be less likely to intervene in a meaningful way if legislation has already provided an appeals process for decisions.[338] As most claims will concern the exercise of a decision-making power provided under statute, the extent to which the decision-maker's role is expressly spelt out – including the matters that must or may be taken into account – will also affect the viability of a claim.

---

332 *Hamilton City Council v Waikato Electricity Authority* [1994] 1 NZLR 741 (HC) at 757.

333 *XY v Attorney General* [2016] NZHC 1196, [2016] NZAR 875 at [60].

334 Cooke, above n 326, at 18.

335 See Grant Illingworth QC "Discretion, Legality and the Bill of Rights" (paper presented to the New Zealand Law "Society Using Human Rights Law in Litigation" Intensive Conference, June 2014) 1. See also: *Pharmaceutical Management Agency Ltd v Roussel Uclaf Australia Pty Ltd* [1998] NZAR 58 (CA) at 66 per Blanchard J; *Ye v Minister of Immigration* [2009] 2 NZLR 596 (CA) at [303] per Glazebrook J; *Waitakere City Council v Lovelock* [1997] 2 NZLR 385 (CA) at 403 per Thomas J; *Mihos v Attorney-General* [2008] NZAR 177 (HC) at [101]; and *Taylor v Chief Executive of the Department of Corrections* [2015] NZCA 477, [2015] NZAR 1648 at [89].

336 Matthew Smith *New Zealand Judicial Review Handbook* (2nd ed, Thomson Reuters New Zealand Ltd, Wellington, 2016) at 530 - 534.

337 Michael Taggart "Proportionality, Deference, Wednesbury" (2008) NZ L Rev 423. See also Dean R Knight "Mapping the Rainbow of Review: Recognising Variable Intensity" (2010) NZ L Rev 393; and (for a more sceptical view of court practice) Claudia Geiringer "Sources of Resistance to Proportionality Review of Administrative Power under the New Zealand Bill of Rights Act" (2013) 11 NZJPIL 123.

338 See *Tannadyce Investments*, above n 325.

C Grounds of review

Assuming an algorithmic decision is reviewable, an applicant for review could (depending on the circumstances) rely on a range of grounds to show that the algorithmic use is unlawful.

1 Improper taking account of irrelevant considerations

PSAs may be particularly vulnerable to a claim that a decision-maker improperly took into account irrelevant considerations.[339] To show that a decision-maker has taken into account irrelevant considerations, the applicant will need to show that a factor was irrelevant to the empowering provision (or common law power), and was material to the decision "in the sense of actually influencing it to be taken".[340] Relevant to this question will be whether the decision-making power requires the taking into account of a closed list of factors, or whether the decision-maker may (expressly or by implication) consider a non-stated criterion.[341]

Algorithmically directed decisions are particularly open to challenge for irrelevant considerations. For, example an algorithm relied upon by a decision-maker could effectively automate a decision. If this automated decision relies on a large range of information (e.g., because this broader sweep of information correlates with the highest degree of accuracy in decision-making), the decision could be unlawful to the extent it has taken into account information falling outside of a closed list of decision-making criteria.

Directed decisions based on open-ended criteria might also be vulnerable. In these cases the relevance of any matter taken into account will be highly contextual and will likely turn on the "text and purpose of the empowering provision".[342] However, if the algorithm relies on factors which, while shown to have some correlation, on their face appear irrelevant, an applicant may also be able to make a claim based on irrelevant considerations. For example, suppose an algorithm was used to help determine whether someone should receive ACC funding for an injury, and it found a correlation based on a person's income. Although this information might be correlative (perhaps those on high incomes are less

---

339 *Poamanga v State Sevices Commission* [1985] 2 NZLR 385 (CA).

340 Taylor, above n 227, at 814.

341 At 804.

342 At 814.

likely to make false ACC claims), it would likely be arbitrary and irrelevant to factor in income to make this decision.

However, several further points are worth exploring. First, the importance of materiality described in chapter four for HRA claims will be equally relevant to a claim that irrelevant considerations have been taken into account. In most cases it will be important to operate a two layer test:

(1) Have irrelevancies been taken into account by the algorithm and, if so, are these material to the algorithm's overall output?

(2) If irrelevancies are material to the algorithm's output, has the output been relied upon such that the irrelevancies are material in the sense of actually influencing the decision?

The above process again illustrates how directed decisions are more vulnerable to review because, provided the first step is met, there will almost always be material reliance on the algorithm's output. But in the case of informed decisions – at least for open-ended discretions – the matter will turn on the degree to which the decision-maker was actually influenced, given other factors also taken into account, and/or understood the algorithm's limitations.[343] The discussion of *Loomis* above, concerning prisoner sentencing tools, makes this point clear – although if that case occurred in New Zealand, it seems likely the judge's use of an unreliable tool would meet the irrelevancies threshold, even if other considerations were important.

Second, in relation to open-ended decision-making discretions, the accuracy, validity and reliability of the algorithm may influence what is "relevant". Suppose an algorithm has a vast number of criteria or decision-points, some of which would normally be considered irrelevant (e.g., a person's income as described above). Arguably, evidence that these criteria improve the algorithm's outputs *is also evidence of their relevance*. This is the very

---

[343] Oswald notes that to avoid inappropriate use, a decision-maker relying on an algorithmic output must "determine whether the decision under consideration matches the one for which the algorithm was developed—for instance, an assessment of 'risk' may encompass much more than the forecast of a particular behaviour by an algorithm—and whether the data on which the algorithm was trained match the circumstances of the current situation". See Marion Oswald "Algorithm-assisted Decision-making in the Public Sector: Framing the Issues using Administrative Law Rules Governing Discretionary Power" (2018) 376: 20170359 Phil Trans R Soc A 1 at 7.

power of algorithms, and ML algorithms in particular: the ability to show the relevance of connections and correlations that might otherwise be hidden or appear nonsensical.

This view naturally follows if one agrees that standards of relevance applied to algorithms need not exceed those that apply to human decision-makers. If an algorithm is valid and generally highly accurate and reliable – perhaps much more than a human – then the fact that some otherwise "irrelevant" considerations are used by the algorithm may not matter. Zerilli et al, for example, warn against imposing a double-standard between humans and machines.[344] On the other hand, the discussion in chapter two highlights that accuracy is not everything in decision-making. Other concepts of "fairness" remain relevant, as do normative principles about how decisions are made.

Moreover, the "accuracy" argument hits up against a reoccurring question in this thesis: accurate for whom? An algorithm's general accuracy across a statistical dataset does not mean that it will be accurate when applied on an *individualised basis* to the specific circumstances of the person about whom a decision is being made. As described in chapter two, the assumptions that apply for a dominant demographic group may not always apply accurately to an "outlier" demographic group to which the individual belongs. Moreover, the decision-making criteria may expressly or impliedly require close examination of the particular characteristics of the individual concerned, and not what others in a similar situation tend to be like. As such, merely pointing to the accuracy for a general group may not be sufficient for a PSA to defend the relevance of an reliance on an algorithmic tool, if it is not well adjusted for the individual(s) affected.

2 Failure to take account of mandatory relevant considerations

An applicant could also show that a decision-maker has also failed to take account of mandatory relevant considerations when using an algorithm. Mandatory considerations may be specified in a closed list, or may arise from the context of the decision.[345] Where mandatory considerations are not set out, *CREEDNZ Inc v Governor-General* states one should consider whether the scheme "expressly or impliedly identifies considerations required to be taken into account… as a matter of legal obligation", and that "the more general and more obviously important the consideration, the readier the Court must be to

344 Zerilli et al, above n 233.

345 Cooke, above n 326, at 29.

hold that Parliament must have meant it to be taken into account."[346]

A failure to take account of mandatory relevant considerations is likely to arise in three scenarios. First, this may occur where there is a directed decision (i.e., the algorithm is entirely or primarily relied upon by the decision-maker) and the algorithm has been constructed in such a way that it does not "cover" the mandatory considerations. Provided there is evidence of what the algorithm *does not do*, it should be relatively easy to prove a failure to consider mandatory matters where there is a closed list of considerations. The question will naturally be harder where there is an open list because there will always be room to argue what is actually mandatory as opposed to a permissible consideration.

Second, even where the algorithm does cover the mandatory grounds of review, it is a truism that "a decision maker must give genuine, and not merely token or superficial regard, to mandatory considerations".[347] However, this is a real risk with algorithmic decision-making. In particular, a decision-maker's reliance on automated tools can lead to the decision-maker failing to give "genuine attention and thought"[348] to a mandatory criterion. The studies of airline pilots who use automated systems attest to the danger that, when a tool is designed to respond to a certain matters, one is less likely to evaluate those matters independently or as deeply and is more likely to reallocate mental load to other areas.[349] In most cases, even if the algorithm does speak to some mandatory relevant considerations, this kind of outsourcing of genuine consideration will be unlawful. And even where the decision-maker shows genuine thought, in order to satisfy his or her decision-making role, the decision-maker arguably needs to understand the limits of the algorithm and its levels of accuracy and reliability.

This leads on to the third possible scenario: where the decision-maker's use of the algorithm means that other mandatory considerations, lying outside of the competence of the algorithm, are not properly taken into account in an informed decision. Consider again the benefit surveillance scenario. If a MSD staff-member put too much weight on a "red" notification alone, the staff-member might not consider other matters which are either expressly or implied required. Important information – say that the individual was already closely monitored and frontline staff had ruled out any issues – might be ignored.[350]

---

346 *CREEDNZ Inc v Governor-General* [1981] 1 NZLR 172 (CA) at 183.

347 *Ye v Minister of Immigration*, above n 335, at [90].

348 *New Zealand Fishing Industry* Association, above n 324, at 551.

349 See Onnasch et al, above n 87.

350 See the discussion of *Belcher v Chief Executive of the Department of Corrections* [2007] 1 NZLR 507

Essentially, the over-reliance on automation as described above, can lead to tunnel vision and a failure to scan for other considerations which may be mandatory.

3 Abdication and self-fettering of discretion

For essentially the same reasons as outlined above, an applicant may be able to show that a decision-maker has abdicated his/her role or impermissibly fettered his/her own decision-making discretion. In particular, the dangers of automation complacency and the search for ever increased efficiency in the use of public resources create a heightened risk that decision-makers will improperly let an algorithm assume control of their decision-making role.

Decision-makers granted a discretion to make a decision must "not abdicate the authority" to do so,[351] nor "preclude [themselves] from inquiring into matters which are relevant".[352] A decision-maker exercising a statutory discretion also must not "shut his ears" to new information.[353]

While a PSA can set a policy to encourage consistency, it must allow space for the proper exercise of discretion; "reliance on policy is not unlawful, but blind following of policy is".[354] Where an algorithm effectively enforces a policy through consistent application of its logic, following the algorithm's recommended action is not *per se* an abdication – provided always that the decision-maker actually understands how the algorithm has reached its decision, independently considers the decision, and *in practice* does reject it where appropriate.

But, the wholesale adoption of directed decisions will almost always create grounds to argue that the decision-maker has wrongly fettered or abdicated his/her decision-making role. When a policy is applied through automated means, the decision-maker risks failing to give real thought to the decision and whether the algorithm's recommendation should be overridden. For this reason, the Australian Government has suggested that while it is

---

(CA) below.

351 *Attorney-General v Unitec Institute of Technology* [2007] 1 NZLR 750 (CA) at [29].

352 *Broadcasting Corp of New Zealand v Broadcasting Tribunal* [1986] 2 NZLR 620 (CA) at 634.

353 See *British Oxygen Co Ltd v Minister of Technology* [1971] AC 610 (HL).

354 *Criminal Bar Association of New Zealand Incorporated v Attorney-General* [2013] NZCA 176 at [118] - [119]. See also *Westhaven Shellfish Ltd v Chief Executive of Ministry of Fisheries* [2002] 2 NZLR 158 (CA) at [45].

possible to make decisions automatically, "the authority for making such decisions will only be beyond doubt if specifically enabled by legislation".[355] Likewise, in the English context Le Sueur believes "the prudent conclusion" is that express legislative authority is needed to empower automated decision-making with little human involvement.[356]

As outlined in chapter two, in New Zealand the recently passed Court Matters Act 2018 is the only notable example where clear statutory provision has been made for automated decision-making. However, Le Sueur has speculated that the English "RAM doctrine" might apply in some cases to legitimise the ceding of decision-making authority to algorithms.[357] That doctrine is analogous to New Zealand's "third source", which suggests government departments may undertake the same actions as natural persons without additional legislative authority.[358] But reliance on the "third source" is unlikely to protect a PSA where statutory decision-making powers have already been defined, and have been fettered or abdicated through the use of an algorithm. This leaves PSAs vulnerable to the extent they rely on algorithmically-generated results and tend to follow these without clearly recording how discretion has been exercised.

As with irrelevant considerations, the applicant's chance of success under this ground will be lower in the case of informed decisions. Here decision-makers can maintain that they independently made a decision, even if they had regard to the algorithm's results or recommendations. But, these assertions are still vulnerable if there is no contemporaneous record indicating the independent exercise of discretion; in such cases an applicant can argue that the decision-maker operated in "auto-pilot" and has cursorily adopted the algorithm's suggestions. On the other hand, success will likely turn on the courts' assessment of the evidence – and a court may well take decision-makers at their word absent compelling evidence to the contrary. As such, a court may be reluctant to suggest someone has fettered his/her discretion, unless the decision-maker's veracity or credibility has been challenged or there is a failure to show clear individualised consideration of the factors relevant to the affected individual.

---

355 Australian Government, above n 184, at 35.

356 Andrew Le Sueur "Robot Government: Automated Decision-making and its Implications for Parliament" in Alexander Horne and Andrew Le Sueur (eds) *Parliament: Legislation and Accountability* (Hart Publishing, Oxford, 2016) 183 at 195.

357 At 194.

358 See the Supreme Court's discussion of the "third source" in *Quake Outcasts v Minister for Canterbury Earthquake Recovery* [2015] NZSC 27, [2016] 1 NZLR 1 at [109] - [121].

It is also worth noting that some algorithms can enhance decision-making and the exercise of discretion, provided they are carefully created and used so as to support the decision-maker's role. For example:[359]

> When properly designed and modelled, automated systems may enhance the exercise of discretion by the following measures:
>
> - Only permitting the use of human discretion and judgement where it is relevant
> - Outlining and/or breaking down the factors decision-makers should consider when making their judgement
> - Providing links to relevant support materials and guides
> - Requiring that the decision-makers clearly state and record reasons for decisions, as a statement of reasons or other official (and auditable) output.

Bearing this in mind, how the algorithm is intended to function and assist the decision-maker – and whether this is more or less likely to reduce independent exercise of a discretion – will be highly relevant to the chance of an applicant being successful. An algorithm that acts to flag relevant information and prompt for things to consider, may point the other way.

4 Improper delegation

An applicant could also argue that there has been impermissible delegation of decision-making powers to an algorithm. The argument here is essentially the same as for abdication – a decision-maker who automates a decision or always follows an algorithm's recommendations for a decision, effectively "delegates" his or her decision-making powers to the machine. Authors such as Le Sueur and Oswald have suggested that, in the English context at least, this could provide another ground to attack a decision.[360] Gavaghan et al have also suggested that it could apply in New Zealand.[361]

However, it is hard to see that this ground adds anything beyond that described for self-fettering or abdication. Also, arguably there is no "delegation", because delegation requires delegation to another legal actor (and an algorithm has no independent legal personality). The argument might be more tenable where the algorithm is overseen or provided by a

---

359 Australian Government, above n 184, at 14.

360 Le Sueur, above n 356, at 190 - 195; Oswald, above n 343 at 14.

361 Gavaghan et al, above n 81, at 40.

third party. In that case, an applicant *might* be able to claim the third party has been delegated the decision-making responsibility.

5 Material error of fact

An applicant may also be able to prove a decision-maker's reliance on an algorithm has caused a material error of fact. This can occur if the decision-maker is "led into mistake and [fails] to take into account true facts" due to the algorithm.[362] In substance, this ground is likely to have a high degree of overlap with arguments that irrelevant considerations have been taken into account (noting also, that irrelevant considerations need to have a material impact to lead to an unlawful decision).

There are at least three ways in which a material error of fact may occur in relation to an algorithmic decision. First, the algorithm may rely on data which is old or inaccurate, so as to produce predictions or outputs that are incorrect or unreliable (the so-called "dirty data" problem). Second, the algorithmic model itself may be constructed in such a way that it is biased or inaccurate, either generally or when applied to a particular category of person, even if the data is otherwise accurate. For example, this may occur if the algorithm was only validated on one demographic group, and is applied to a different group. Third, error of fact can occur where the reliance on the model alone or without proper regard to other contextual factors leads a decision-maker to disregard information that disqualifies the algorithm's output. As Gavaghan et al have noted,[363] the Court of Appeal's comments in *Belcher v Chief Executive of the Department of Corrections* neatly summarise the danger of relying on an algorithmic output (in this case, derived from the Department of Corrections' ROC*ROI tool) without regard to other factors which will impact its accuracy:[364]

> Obviously factors which have arisen post-release must be allowed for in an ESO assessment. For instance, if the appellant had been rendered a tetraplegic as a result of a post-release accident, this would have presumably eliminated the likelihood of him re-offending and would undoubtedly have negated any adverse inferences which might otherwise have been drawn for actuarial assessment.

---

362 *Daganayasi v Minister of Immigration* [1980] 2 NZLR 130 (CA) at 149.

363 Gavaghan et al, above n 81, at 53.

364 *Belcher v Chief Executive of the Department of Corrections*, above n 350, at [90].

Although an appeal under the Parole Act 2002 rather than a judicial review case, *Belcher* is also instructive of the need for these issues to cause a material impact on the decision. In that case the appellant had been convicted of child sex offences and was subject to an Extended Supervision Order ("**ESO**") that would impose intensive monitoring after his prison term. The Court acknowledged that the application of the ROC*ROI to the appellant's circumstances was problematic because:[365]

> … it was not designed to predict risks of recidivism in relation to those who had been in the community for some years. For this reason, the measures did not take into account the period of time which the appellant had spent in the community since his release from prison. In turn, this has the consequence that the actuarial assessments of the appellant's risk of recidivism could only fairly be applied to the appellant if appropriate allowance was made for the time which he had spent in the community without further sexual offending.

Despite this evidence, the Court did not disturb the imposition of the ESO, as a lack of offending was not determinative given the typically sporadic and opportunistic nature of child sex offending and because other assessments had supported the imposition of the ESO.[366] A court could well take a similar view in a judicial review case, even if the algorithm is inaccurate. Nevertheless, subject to the extent to which errors caused through the algorithm are "material" to the final decision, review on the basis of material error of fact responds to many of the perennial issues with algorithms described in chapter two.

6 Procedural unfairness and the right to reasons

The discussion above has shown how algorithms can contribute to unlawful decisions by affecting the information relied upon by, and the role of, the relevant decision-maker. However, an applicant may also be able to show that a decision-maker's use of an algorithm caused procedural unfairness, making the decision unlawful.

Procedural fairness (or natural justice) is based on two key principles: that "the parties be given adequate notice and opportunity to be heard… and that the decision maker be disinterested and unbiased."[367] The expectation that public bodies will conduct themselves

---

365 At [88].

366 At [88] - [93].

367 *Combined Beneficiaries Union Inc v Auckland City COGS Committee* [2008] NZCA 423, [2009] 2 NZLR 56 at [11].

in accordance with the requirements of natural justice is fortified by s 27(1) of NZBORA:[368]

> **27 Right to Justice**
>
> (1) Every person has the right to the observance of the principles of natural justice by any tribunal or other public authority which has the power to make a determination in respect of that person's rights, obligations, or interests protected or recognised by law.

What is necessary for natural justice or procedural fairness will vary significantly depending on the nature of the impact on the individual(s) and the context of the decision-making power.[369] Legislation will often define the parameters (or preclude) aspects of natural justice in a given context.[370] Any appeal rights will also be relevant. Moreover, the nature of the interests affected by a decision will be particularly relevant to the procedural hygiene necessary from a decision-maker.[371]

In relation to algorithmic decisions, a failure to be heard is particularly likely to be relevant. First, while context will always affect matters – a court is unlikely to require an affected party's input where an algorithm makes an automated tax refund for Inland Revenue – the risk is that use of an algorithm narrows the focus of the decision-maker (as discussed above). In some cases the decision-maker may fail to give individuals the opportunity to comment that they deserve.

Second, where a decision-maker *does* need to take into account an affected party's views, information that could substantially disturb the algorithm's output might be ignored. This is particularly likely where a decision-maker "rubber-stamps" algorithmic recommendations, but is also possible where the decision-maker has insufficient understanding of the algorithmic system to be able to be assess whether the algorithm's results should be disregarded. In this case, there can be an argument that the decision-maker has failed to hear the party.

Third, in those cases where a right to be heard exists, arguably so does a right for the

---

368 At [50].

369 *Combined Beneficiaries Union*, above n 367, at [11]; and Francis Cooke "Judicial Review" (New Zealand Law Society seminar, May 2012) at 33.

370 Cooke, above n 326, at 33.

371 *Combined Beneficiaries Union*, above n 367, at [11].

affected person to receive a proper explanation of an algorithm's workings, so as to be able to exercise that right. For example, in the context of high-stakes decisions related to criminal justice (e.g., sentencing or charging) or immigration (e.g., deportation or asylum claims), the right to voice is meaningless if the affected party cannot receive an adequate explanation of how the algorithm generally produces its outputs, and how it has been applied specifically to the party. This knowledge allows the party to test and influence the decision-maker. And yet, as discussed above, ML algorithms can be incredibly hard to interrogate, potentially subverting the party's rights. In such high-stakes cases, there is the real prospect that a PSA's failure to provide an adequate explanation of the algorithm would be unlawful.

Separately, decision-makers can also be subject to an obligation to provide reasons at the same time as or after making their decision. This obligation can arise under statute or common law.[372] Relevantly, a failure to give sufficient reasons – depending on what the context requires – can be a reviewable error of law.[373] This could be particularly important to the extent that judicial or quasi-judicial bodies do not fully demonstrate that they understand how an algorithm contributed to a decision.

As chapter three highlighted, a PSA will usually be required to give reasons where this is requested under the OIA.[374] While this is not a proactive obligation, it is broad reaching – meaning that a motivated party that could use it as grounds to initiate judicial review in most cases, if the reasons provided did not meet the level of disclosure required by the OIA. However, as discussed above, the disclosure requirements under the OIA may actually be quite limited, and will likely fall well short of a full technical explanation of how an algorithm creates a particular output.

The right to adequate reasons, and the ability to seek review where this is not met, is likely to be more relevant instead in a judicial or quasi-judicial context – for example, where a sentencing judge or the Parole Board make reference to an algorithmically-generated prediction of an individual's risk. The common law is likely to provide a right to reasons in these cases, absent a specific statutory scheme outlining what must be provided.[375] The right to reasons in these setting is particularly important to ensure openness in the

---

372 Taylor, above n 227, at 313.

373 See *Lewis v Wilson & Horton* [2000] 3 NZLR 546 (CA) at [86] - [87].

374 OIA, s 23; LGOIMA, s 22.

375 *Lewis v Wilson & Horton,* above n 373.

administration of justice, to ensure lawfulness can be assessed by the High Court exercising its supervisory jurisdiction (i.e., in judicial review proceedings), and to ensure a "discipline" for decision-makers that protects against "wrong or arbitrary decisions and inconsistent delivery of justice".[376]

What is required for reasons to be adequate in judicial or quasi-judicial setting? In *Television New Zealand Ltd v West* the High Court outlined how the "depth of the reasoning process can be expected to vary in accordance with the role of the tribunal and the nature of the hearing".[377] However, Durie J's comments in *Re Vixen Digital* suggest that decisions: (a) "must be sufficient to enable anybody with a power of review to understand the process of thought whereby a conclusion was reached"; (b) must allow those who interests it affect "to so understand the basis for decisions as to be better informed in predicting that which is or is not within the law"; and (c) where there is some general public interest in the decision, should allow the public to know and comprehend the standards the decision-maker sees as important.[378] The upshot of this, especially in areas that concern important human rights, is that a tribunal may need to show in its reasoning that it understands the limitations of an algorithm, show how the algorithm contributed to its decision, and (possibly) provide an explanation of the algorithm's technical working to the affected party.

D Judicial review: practical considerations

Lastly, even if an applicant has strong grounds to review a decision, it is important to consider how practicalities affect the potential of judicial review to respond to an algorithmic decision.

First, a claim for judicial review can be expensive. Unlike a HRA or Privacy Act claim which can be undertaken by the Director, affected parties will have to fund the proceedings on their own. Moreover, like a HRA claim, the evidence provided by the PSA can be extensive and require significant resources to review. In short, parties are unlikely to bring a High Court proceeding in judicial review unless there are very good reasons – for example, because they have been badly affected by a decision. Therefore, algorithmic decisions which only lead to low level harm (or which affect poorer parties) are less likely

---

376 At [76] - [85].

377 *Television New Zealand Ltd v West* [2011] 3 NZLR 825 (HC) at [82].

378 *Re Vixen Digital* [2003] NZAR 418 (HC) at [43].

to be reviewed.

Second, judicial review provides very different remedies from the Privacy Act or HRA. Remedies are inherently at the discretion of the Court, although there must be "extremely strong reasons to decline to grant relief".[379] Moreover, unlike the HRA or Privacy Act jurisdictions, the Court will not grant damages. Instead, the focus is typically on quashing the decision, stopping any ongoing unlawful practice and/or having a decision re-heard in a lawful way.[380] Even with a costs judgment in the applicant's favour (which is not assured), these remedies will not compensate a party for any loss which has arisen from a decision, let alone the burden of taking proceedings.

However, the remedies provided by judicial review can be significant, particularly for high stakes cases or where the purpose is to prevent an ongoing practice adverse to a group. For an individual who is set to be deported, a successful claim on the basis of a breach of natural justice will be incredibly significant, as the usual remedy is to have a decision quashed.[381] Even if the result is that the decision-maker will reconsider the case without reference to an algorithmic output, this will be a new second chance. Moreover, in other cases, an order that the decision-maker do or refrain from doing something in relation to the use of an algorithm – for example, this could concern how DHBs use an algorithm to prioritise medical treatment[382] – could have a wider public benefit to those who might be affected by potential algorithmic harms. In this way, a judicial review proceeding can have an ongoing prophylactic impact on public sector behaviour and encourage political accountability, particularly where subject to media coverage.

E Judicial review: conclusion

The above discussion therefore shows that judicial review is potentially the most effective tool to challenge a particular algorithmic decision or the practice of using algorithms in a

---

379 *Air Nelson Ltd v Minister of Transport*, above n 329, at [60].

380 See Judicial Review Procedure Act 2016, ss 16 - 18.

381 A decision will usually be quashed unless this would "bring about unacceptable administrative consequences of inequity to third parties". See *Air Nelson Ltd v Minister of Transport*, above n 329, at [74]. However, it is worth noting that the availability of judicial review proceedings is limited under Part 7 of the Immigration Act 2009 and will depend on the context.

382 See Rashmi Dayalu, Elizabeth T Cafiero-Fonseca, Victoria Y Fan, Heather Schofield and David E Bloom "Priority Setting in Health: Development and Application of a Multi-criteria Algorithm for the Population of New Zealand's Waikato Region" (2018) 16 Cost Eff Resour Alloc 35.

given context, given the wide range of ways in which a PSA can fall foul of administrative law principles. However, there are a number of points to keep in mind, which will affect the ability of judicial review to respond to algorithmic harms.

Like a HRA claim, the line between directed and informed decisions will often be critical. If a decision is largely or completely automated, this should make it easier for a person to claim that there has been unlawful fettering of decision-making discretion – and PSAs should be extremely wary of this kind of automation without explicit empowering provisions. Automation will also improve the chances of a claim based on irrelevant considerations. As with the HRA, the contribution of the irrelevancy will need to be material, and this is more likely when the algorithm's input is the sole or major factor in the decision, rather than one of many. As most decisions will not be solely automated, the influence of the algorithm will be a keenly fought evidential matter.

Perhaps most importantly, judicial review claims have a better chance of addressing systemic harm caused by government practice than claims based on the Privacy Act and the HRA. This is because a judicial review claim brings the possibility that a particular practice is held unlawful, and the agency will need to adjust its practices – without the need to prove harm.

However, the remedies available under judicial review still have mixed utility. While judicial review can encourage forward-facing public sector accountability, it will not compensate those affected by an algorithmic decision for past harms. On the other hand, judicial review can be just as meaningful in high stakes environments – such as criminal justice or asylum decisions – where the quashing of a decision can protect an individual's physical liberty or other rights.

Lastly, the costs of judicial review means that it is only likely to be taken in rare cases – limiting its true value as a counterbalance to poor algorithmic practice.

Having now explored the responsiveness of judicial review and other relevant legal frameworks, the next chapter of this thesis proposes a regulatory model that legitimates proportional use of algorithms, while limiting remaining potential harms.

VI Chapter Six: A New Regulatory Model for New Zealand

A Overview

The previous two parts of this thesis have traversed the growing use of algorithms in New Zealand and overseas, and closely examined the extent to which New Zealand's existing laws could provide recourse for those adversely affected by algorithmic decisions. This final part gathers these threads together to argue that, even with tweaks to existing legal protections, a carefully constructed regulatory model is needed to ensure a proportional balance between the benefits of algorithms and the risks of harm.

First, we explore the gaps in the patchwork of existing legal protections to contextualise why a regulatory model is needed.

B Protections from algorithmic harm: the gaps in the patchwork

As outlined in chapter two, the use of algorithms can create a range of risks. Harms can arise from self-justifying feedback loops, the systemic classification of people, and inferences that overcome the purpose of privacy protections. Harms can also arise when individuals cannot access why or how a decision has been made, and when decision-makers treat algorithms as objective or independent actors. Demographic outliers are often particularly affected by algorithmic harms. In practical terms, these harms can be direct and immediate (e.g., the imposition of a harsher sentence) or cumulative and gradual (e.g., where surveillance slowly changes one's behaviour).

The discussion in previous chapters highlights how existing laws have typically been crafted according to policy goals which in principle address these algorithmic harms. These goals include a presumption that the state has limited rights to intrude into individuals' private affairs, and that decisions should be made transparently, based on accurate and relevant information, via a fair process and without discriminating between similarly placed groups.

The proposed changes to existing laws – particularly the Privacy Act, OIA and the regime for unreasonable search and seizure – will go some way to improving the responsiveness of existing legal avenues. Tweaks can ensure algorithmic outputs and "passive" surveillance activities are subject to the Privacy Act's protections. Likewise, transparency and fairness in decision-making would be enhanced by new rights under the OIA for

individuals to receive notice of an algorithmic decision and to access human review. An existing "right to reasons" could be adjusted to the contours of algorithmic decision-making. And changes could ensure reasonable expectations of privacy can arise in public places, in relation to protections against unreasonable search and seizure.

Nevertheless, gaps remain. The first of these should be clear: access to a legal remedy will often hinge on the line between directed decisions and those which are merely informed by the use of an algorithm. Generally this is because a directed decision helps to establish the causal relationship between the algorithm and any related harm. However, the borderline between a directed decision and an informed decision will not always be clear, and neither will the harm from an informed decision necessarily be less – it just may be harder to prove. So, situations where algorithms *do* contribute to real harm, despite occurring through an informed decision, will be hard to challenge. It may be that these cases are rare. However, without further empirical research, there is nothing discrediting this potentially significant gap.

Second, the remedies available under existing protections only provide limited compensatory recourse. For example, the Privacy Act may only provide meagre monetary damages for individuals affected by an interference with privacy. While the HRA might provide more significant monetary compensation, the legal threshold to bringing a case is high. Even if these avenues provide monetary recourse in clear-cut cases, as indicated above, this may be difficult to access in the "grey" cases where the influence of the algorithm is harder to substantiate. Judicial review, meanwhile, will not provide any monetary remedy, and the right to be free from unreasonable search and seizure will only provide monetary remedies in the narrowest of cases.

This leads to a very important third point: these legal avenues will often fail to respond to lower-level harms, even if these risk accumulating and causing other impacts at a later time. The rights-based nature of the HRA and Privacy Act focus on substantial harms to individuals, and do not easily address more diffuse harms across groups. Although the Privacy Act does provide a limited power for the Privacy Commissioner to inquire into agencies' conduct, this function is without remedial powers and is necessarily reactive rather than preventative.[383]

Fourth, judicial review provides an easier legal avenue than the HRA or Privacy Act, and

---

383 Section 13(1)(m).

one that *can* actually change PSA practices which cause these lower level harms. However, judicial review is still a reactive remedy which cannot compensate for harm done.

Lastly, all of the remedies are affected by the cost and time necessary to bring a case. Although judicial review brings the greatest potential for adjusting PSA practice, it is expensive and likely to be used only by the most motivated parties. And while claims under the HRA and Privacy Act can be supported by the Director – but will not always be – access to justice still entails a longwinded and painful process.

In sum, these legal avenues provide a useful but limited framework to deter and/or compensate for algorithmic harms. In particular: informed decisions are less likely to be successful (whether or not they lead to significant harm); only motivated and/or well-resourced individuals are likely to take claims; the adequacy of recourse will turn significantly on the context; and group-based cumulative harms are not easily addressed (although judicial review can influence ongoing PSA practice). Perhaps most importantly, these remedies are the ambulances at the bottom of the cliff – they do not ensure PSAs implement sensible processes that balance the benefits and potential harm from using algorithms. A broader framework is necessary to ensure adequate protection.

C Rationale for a new regulatory model

Given the gaps mentioned above, this thesis calls for a new regulatory model for public sector use of algorithms. There are several reasons why this model is needed.

First, a new model can address potential harms to the public which are not easily accommodated by rights-based claims or through judicial review. In particular, a regulatory model can help reduce the likelihood of harms arising in the first place by requiring agencies to evaluate their decision-making practices, and by providing a process for independent review. This top-down approach creates a better chance of addressing some of the more diffuse harms that can arise with algorithms – such as place-based algorithmic policing, and harms arising from the allocation of resources above the individual level.

One may ask: are the remaining potential harms, after tweaks to the informational rights outlined above, really sufficient to justify a new regulatory model? In short, yes. As outlined above, even when there are substantial harms, practical barriers can prevent access to existing remedies, limiting their effectiveness.

This thesis suggests that algorithmic harms should be thought about in a similar way to

environmental pollution.[384] For example, even if one does not become ill after swimming in a polluted river, it does not mean that this risk did not exist or that poor environmental practices upstream should be ignored. Likewise, one should remain concerned about a PSA's poor algorithmic hygiene, whether or not this produces immediate harm that is material enough to establish a Privacy Act or HRA claim.[385] The difficultly in directly quantifying this risk of harm does not necessarily justify delaying action. As Pasquale notes:[386]

> Just as we cannot quantify in monetary terms all forms of human transformation of the natural world that are discomfiting enough to merit legal regulation, we will not always be able to offer precise valuations of the alarm or apprehension we feel at certain algorithmic transformations of human social relations.

Support for a new regulatory model is also found in the contention that a high standard should be applied when departing from the long-standing and well-understood status quo of human decision-making – particularly given the state's unique ability to exercise legitimate coercive power. As highlighted in chapter three (in relation to a right to human review), this thesis argues that a soft form of the "precautionary principle" provides a sound starting point when thinking about algorithmic harms.[387] While the precautionary principle has been interpreted in various soft and hard forms,[388] at its core it argues that in the face of scientific uncertainty, decisions which could cause ongoing damage – or at least damage for which money fails to adequately compensate – need to be justified in advance.[389]

Aligning with this principle, the proposed regulatory model's use of algorithmic impact assessments or AIAs (described below) is intended to ensure that the potential dangers from algorithmic use are understood and justified before an algorithm is deployed. This is consistent with the rationale for data protection impact assessments required under the

---

384 Other commentators similarly see this parallel. See Andrew Tutt "An FDA for Algorithms" (2017) 69 Admin L Rev 83; and Balkin, above n 11, at 1232 - 1235. While generally open to the environmental pollution analogy, Pasquale suggests there remains a distinction between environmental harm, which can be measured on a scientific basis (e.g., x percent of air pollution), and the less quantifiable harms which can arise through algorithms. See Pasquale, above n 98, at 1251.

385 See Gandy, above n 66, at 37 - 38.

386 See Pasquale, above n 98, at 1251.

387 For an explanation, see Kriebel et al, above n 239.

388 See New Zealand Treasury *Environmental Risk Management in New Zealand: Is there a Scope to Apply a More Generic Framework? – Policy Perspectives Paper* (Wellington, July 2006) at 12 - 13.

389 Luiz Costa "Privacy and the Precautionary Principle" (2012) 28 Comp Law & Sec Rev 14 at 23 - 24.

GDPR in certain cases.[390]

Moreover, an appropriate regulatory model could create a mechanism for political (rather than strictly legal) accountability for policy decisions involving challenging trade-offs between groups. Parliament and the public could judge how algorithms are used and the PSA's (and/or Minister's) justification for any collateral harms. This would encourage accountability at a broader level.

Separately, a well-crafted regulatory model could create a framework which provides proportionality by allowing PSAs to use algorithms for low risk uses, without the prospect of judicial review for an impermissible abdication of decision-making powers. It might also inoculate challenges on the basis of IPP 8 of the Privacy Act, where a considered public policy trade-off could otherwise indicate a failure to take reasonable "reasonable steps to ensure the accuracy" of an algorithm's outputs.

Bearing these matters in mind, this thesis proposes a range of outcomes for a new regulatory model, including:

- proportionate oversight of PSAs' use of algorithms (having regard to the costs of such oversight, and the potential benefits, harms and overall significance of algorithmic uses);
- protection from unacceptable harms, and minimum standards of transparency and procedural fairness, for those who are the subjects of algorithmic decisions;
- processes and protections that ensure citizens' equality of treatment, and PSAs' compliance with the principles of the Treaty of Waitangi, in their use of algorithms;
- political and legal accountability for the circumstances in which algorithms are used in decision-making; and

---

390 See Costa, above n 389; A Naryanan, J Huey and E Felten "A Precautionary Approach to Big Data Privacy" in S Gutwirth, R Leenes and P De Hert (eds) *Data Protection on the Move: Current Developments in ICT and Privacy/Data Protection* (Springer, Berlin, 2016) 357; and Maria Eduarda Goncalves "The Risk-Based Approach under the New EU Data Protection Regulation: a Critical Perspective" (2019) Journal of Risk Research. Pasquale and Citron have also called for AIA type processes in the context of private sector credit scores. See Danielle Keats Citron and Frank A Pasquale "The Scored Society: Due Process for Automated Predictions" (2014) 89 Wash L Rev 1 at 25 - 27.

- within the public sector, the fostering of best practice and broader awareness about the risks and benefits of using algorithms.

Taking these outcomes in mind, the following parts of this chapter outline the form a regulatory model could take.

D Outline of proposed regulatory model

This thesis proposes a regulatory model which combines the role of an independent "Algorithms Watchdog" with a process that ensures algorithms can be used lawfully in administrative decision-making. This model broadly aligns with tentative suggestions put forward by Gavaghan et al in their report *Government Use of Artificial Intelligence in New Zealand*, but also departs in key areas.[391] At a high level, the model would:

(a) use AIAs as a screening and record-keeping tool for proposed use cases for decision-making algorithms;

(b) require potentially risky use cases to be approved by the relevant Minister (including when varied);

(c) legitimise the lawfulness of using algorithms for decision-making (whether in relation to an informed decision or directed decision), provided the algorithm is used only for the purposes described in the use case and subject to any mitigations required by the AIA; and

(d) create transparency requirements by creating a register of decision-making algorithms and their accompanying AIAs.

The key elements of this model are sketched out below. First we turn to options for the "regulator" that will oversee the regulatory model.

---

[391] For example, rather than a suggesting the regulator might have veto powers over controversial algorithmic uses, the approach outlined in this thesis ensures the relevant Minister is politically accountable, and explicitly makes lawful approved uses (thereby limiting the chance that PSAs unwittingly fall foul of administrative law principles). See Colin Gavaghan et al, above n 81, at 71 - 73.

1 Form of regulator

At least four forms are available for the algorithms "regulator". These include: something close to the status quo (i.e., each agency having its own algorithm-related assurance processes, without any external oversight) ("**Self-Regulation**"); a cross-government advocate for best practice, with soft powers to coordinate agencies and issue guidelines (e.g., something similar to the Government Chief Digital Officer)[392] ("**Cross-government Advocate**"); an independent agency that similarly facilitates best practice, but is statutorily independent from Ministerial direction and has algorithm audit and reporting powers to Parliament ("**Independent Monitor**"); or a more hard-edged regulator that, in addition to a role encouraging best practice, also has the power to compel PSAs to change how they are using algorithms and/or to impose pecuniary penalties for non-compliance ("**Hard-edged Regulator**"). This thesis recommends the Independent Monitor approach – encapsulated by the Algorithms Watchdog – as the best fit for the New Zealand context.

Self-Regulation is the least attractive option. This model allows each PSA to oversee its use of algorithms in a way that suits its own purposes. In this sense it provides flexibility. This model also avoids the cost of setting up a separate body with oversight powers. However, there are a range of drawbacks. First, the agency may act in its own self-interest when considering algorithm use cases, potentially at the expense of appropriate consideration of relevant risks. The model also brings the chance of increased cost through duplication of approach, and inconsistencies in process and standards across government generally. The Stocktake Report indicates that this status quo option has created a patchy regulatory approach to the use of algorithms in New Zealand, with a lack of focus on ongoing assurance processes.[393] Inconsistency of approach also creates a greater risk that individuals receive unequal treatment between agencies, and that unsupported agencies inadvertently act unlawfully. Moreover, the level of resource each agency puts into an assurance framework is likely to depend significantly on budget allocation, institutional priorities, and the extent to which senior leadership or Ministers perceive or are aware of the risks posed by algorithms.

A Cross-government Advocate presents an alternative option for soft regulation, but one

---

[392] New Zealand Government "Government Chief Digital Officer" (May 2019) <https://www.digital.govt.nz/digital-government/leadership-and-governance/government-chief-digital-officer-gcdo>.

[393] Statistics New Zealand and Department of Internal Affairs, above n 2.

that has some of the same drawbacks as the Self-Regulation model. A Cross-government Advocate could be put in place without the need for legislative change, probably as part of an existing government department. It would also encourage greater consistency across the public sector, by providing a single office with the authority to issue best practice guidance for how to use and mitigate the risks of algorithms. Its lack of legal enforcement powers could also have benefits – allowing agencies to be frank about where they *are not* following best practice, without fear of formal consequences. This approach is similar to the recommendation that the United Kingdom establish a "ministerial champion" to provide government-wide oversight and coordinate departments' approaches to development and deployment.[394]

On the other hand, locating a Cross-government Advocate within an existing department – the most likely option – creates other challenges. As with the Self-Regulation model, it means the resourcing and independence of the role will likely depend significantly on the department's internal priorities and the extent of Ministerial support – which may change as Ministers and Governments come and go. Further, there is a risk that this body would be too removed from agencies with a high degree of independence from central government. And, its effectiveness would depend significantly on ensuring it had dedicated staff able to influence PSAs' practice through soft skills. Perhaps most importantly, this model would have no formal review and/or auditing role – limiting its ability to provide transparency and political accountability for algorithmic failures within the public sector.

Next we consider the Independent Monitor option. This approach is modelled very loosely on the roles performed by the Office of the Auditor-General[395] and the Parliamentary Commissioner for the Environment,[396] and would be augmented by a legislative framework legitimising agencies' use of algorithms within clear boundaries (described more in the following section). This model is recommended for a number of reasons; most importantly because it achieves a good balance of political and legal accountability without requiring a high degree of hard-edged powers.

First, this body would have separate funding and be statutorily independent from the Government of the day, removing some of the potential obstacles noted in the previous options. Second, it would also play a role in disseminating best practice for the use of

---

394 House of Commons, above n 18, at 3.

395 See Public Audit Act 2001.

396 See Environment Act 1986.

algorithms and, in particular, for AIAs. AIAs – much like privacy impact assessments – would be used to assess any proposed algorithmic use case, the potential harms that arise from this, the practical mitigations that can be put in place (e.g., consultation with potentially affected individuals to help with service design,[397] ongoing audit and validation processes, and internal training) and the remaining residual risk. Under this model, AIAs would be compulsory and become the cornerstone of the regulatory regime.

The Independent Monitor would also put out best practice guidance on common mitigations – such as guidelines for third-party procurement, staff training for algorithmic tools, and for notice to those subject to algorithmic decisions. The body could have powers to coordinate and consult with other relevant agencies and persons, such as the Privacy Commissioner and Human Rights Commission, academics and practitioners with subject-matter expertise, and international counterparts.

The Independent Monitor would ensure accountability in a number of ways. First, it would have the power to request information from, and audit, PSAs in relation to their use of algorithms. This would ensure that agencies put in place the processes outlined in their AIA, and would allow the Independent Monitor to assess the nature of any adverse impacts. This external review could be ad hoc or conducted regularly (e.g., every two years). Second, the Independent Monitor could report annually to Parliament on the use of algorithms in the public sector. This independent reporting would encourage transparency and accountability – allowing opposition parties (or others) to bring problematic behaviour to the attention of the public. Moreover, to the extent a PSA was not acting consistently with the use case recorded in the AIA, this could also create grounds for legal challenge (see below).

While the Independent Monitor would require greater resourcing than the first two options, and its reporting powers could make agencies less likely to volunteer examples of poor practice, it would present a more independent and durable model for reducing the potential harms of the public sector's use of algorithms. This approach is also consistent with the suggestion of Gavaghan et al in their recent report.[398] There the authors also cautiously express a preference for a fully independent agency that would issue best practice guidelines, report annually on use of algorithms and have an ongoing monitoring role

---

397 See AI Now *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability* (April 2018) at 9.

398 See Gavaghan et al, above n 81, at 76 - 77.

across government.[399] Moreover, the practical challenge of setting up a new agency could be substantially addressed by growing this office within the infrastructure of an existing agency. For example, the Office of the Privacy Commissioner could be expanded to include an "Algorithms Commissioner", that would sit within the same building but have independent powers. This would be a natural fit, given the Privacy Commissioner's complementary role regulating the use of personal information generally. Other candidates include the Office of the Ombudsman or the Office of the Auditor-General, given their existing roles scrutinising government.

Lastly, the "Hard-edged Regulator" model would have most of the same characteristics as the Independent Monitor, and could be an alternative option. However, under this model the regulator would also have powers to seek or impose: (a) pecuniary penalties on the PSA for poor conduct; and/or (b) orders to change a PSA's use of algorithms in decision-making.

A Hard-edged Regulator is likely to be disproportionate for regulation of the public sector's use of algorithms. Creating an entity with enforcement powers would further increase resourcing costs, and the prospect of one arm of government taking enforcement action against another arm is generally unattractive. Moreover, enforcement powers are unusual for a regulator of solely government practice, being more typically available to regulators that influence the conduct of private sector actors who are not separately subject to political accountability (e.g., the Commerce Commission and Financial Markets Authority).[400] The transparency provided by auditing of algorithmic use, and reporting to Parliament, is likely to be the main driver of good public sector practice. Lastly, stronger enforcement powers are likely to further disincentivise agencies from self-reporting areas of failure.

2 Algorithmic impact assessments: the regulatory foundation

In addition to an Independent Monitor – encapsulated by the Algorithms Watchdog – the starting point for the proposed regulatory model is the use of AIAs. Before using a new decision-making algorithm, each PSA would be required by statute to complete an AIA to assess the nature of the algorithm, how it would be used, the potential risks involved, and the mitigations that would need to be put in place. As such, AIAs would act as an important screening tool.

---

399 Gavaghan et al, above n 81, at 76 - 77.

400 Financial Markets Authority Act 2011, Part 3; Commerce Act 1986, Part 6.

To ensure a proportionate response, the Algorithms Watchdog could prescribe a two-stage process. Much like the Privacy Commissioner's suggested approach to Privacy Impact Assessments,[401] this would involve an agency first undertaking a short-form AIA to indicate whether a use case had any real risk of harm to individuals. Where no real risks are likely, the short form AIA could adequately record potential issues and why they are not of concern within the bounds of the particular use case. But in most situations, the short-form AIA would be the first step to flush out key issues and allow for a more in-depth analysis of the algorithm and its intended use. This second stage analysis and mitigation process could draw on MSD's PHRaE Framework – which creates a process responding to potential privacy, human rights, and ethical concerns.

AIAs are likely to provide an effective preventative tool to encourage better practice at the system level. As Andrew Tutt notes, it is a characteristic of algorithms that responsibility can be difficult to measure, trace and assign,[402] given the range of actors that can be involved in various stages including design, approval, technical implementation, and operational use.[403] By considering the range of privacy, human rights, transparency and ethical impacts possible through the use a decision-making tool, AIAs would help protect against diffuse algorithmic harms for which individual legal remedies may be unresponsive. AIAs could also take account of the Data Futures Partnership's work on "social licence"[404] – that is, they would also consider the public's comfort levels with the proposed use beyond a purely legal reference point. Each AIA should also consider consistency with the principles of the Treaty of Waitangi, especially given the risk that algorithms do not provide equal levels of validity and/or reliability when applied to minority groups.[405]

The use of AIAs would be consistent with suggestions made by Gavaghan et al[406] and in the Stocktake Report.[407] Moreover, AIAs are similar to other frameworks already used in

---

401 See Office of the Privacy Commissioner *Privacy Impact Assessment Toolkit* (July 2015).

402 Tutt, above n 384, at 105.

403 See House of Lords, above n 170, at 95 - 98.

404 See Data Futures Partnership, above n 167; and Data Futures Partnership and Toi Aria (Massey University) *Our Data, Our Way: What New Zealand People Expect from Guidelines for Data Use and Sharing – Findings from Public Engagement* (March 2017).

405 Observe the Waitangi Tribunal's concerns about the Department of Corrections' use of the RoC*RoI tool in relation to Māori prisoners. See Waitangi Tribunal, above n 6.

406 See Gavaghan et al, above n 81, at 60 - 61, and 73.

407 Statistics New Zealand and Department of Internal Affairs, above n 2, at 33 - 34.

government – such as regulatory impact statements and privacy impact assessments – which substantiate the merits of a proposal by demonstrating the impact of risks and benefits.[408]

There are already a range of useful resources to help influence the requirements for AIAs. Under Canada's *Directive on Automated Decision-Making,* AIAs are used to assess the potential effects of automated decision-making[409] and to identify appropriate levels of transparency, quality assurance, recourse and reporting that may be needed.[410] While the Algorithms Watchdog could craft best practice for using AIAs in New Zealand conditions (following consultation with experts and stakeholders), the Canadian example provides a valuable starting point.[411] Suggestions for how to formulate an AIA process have also been made by AI Now,[412] and AIAs share commonality with the data protection impact assessments or DPIAs required by Art 35 of the GDPR (described above in chapter two). The resources available for DPIAs, and supervisory agencies' practical experiences, could inform the Algorithms Watchdog's guidance.[413]

AIAs would not be needed for every kind of automated tool used by a PSA in a decision. Legislation would need to exclude trivial uses (e.g., business-as-usual software tools like Microsoft's Word or Excel, when used for normal tasks).[414] But this exception should not exclude decision-making which may impact on individuals in significant ways, albeit

---

408 At 33 - 34.

409 Canadian Government, above n 3, at [6.1].

410 At [6]. Note that this covers areas including notice of the use of automated systems before and after use, access to software code, testing and monitoring of outcomes, data quality, peer review, employee training, contingency systems, security controls, access to human review, rights of recourse, and reporting on the effectiveness and efficiency of automated systems.

411 See https://canada-ca.github.io/aia-eia-js/.

412 AI Now, above n 397.

413 See for example, United Kingdom Information Commissioner's Office "Data Protection Impact Assessments"< https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/>; Article 29 Data Protection Working Party *Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing is "Likely to Result in a High Risk" for the purpose of Regulation 2016/679* (4 October 2017, 17/EN WP 248 rev.01); and Alessandro Mantelero "Regulating Big Data: The Guidelines of the Council of Europe in the Context of the European Data Protection Framework" (2017) 33 Comp Law & Sec Rev 584.

414 AI Now, above n 397, at 12.

indirectly.[415] The Algorithms Watchdog could issue guidance on the boundaries between these kinds of tools, and others which would meet the statutory test.

This thesis also suggests a further dividing point based on potential harm after a PSA has gone through the process of putting together an AIA for an algorithm. The test would ask: will the algorithm be used in relation to a decision which could reasonably have a substantial impact on the rights, entitlements, benefits or privileges of the individual (a "**Complex Case**")? If the answer was yes, the agency would need to obtain the approval of the relevant Minister for the use of the algorithm within the use case outlined in the AIA, following consultation with the Algorithms Watchdog.

This process achieves two things. First, it pitches the level of additional oversight to the level of potential harm. In this sense, it is similar to impact-based thresholds in the GDPR (for the use of a DPIA) and in Canada's *Directive*,[416] and also aligns with the suggestion of Gavaghan et al that algorithmic regulation should focus on the possibility of harm, rather than the form of technology.[417] Second, this approach allows political actors to make difficult trade-offs in the use of algorithmic tools – for example, what the acceptable level of "false positives" might be for a tool that flags potential child abusers – which are not easily resolved through the legal process. Because Ministerial decisions on Complex Cases will be made publicly available (see below), it also encourages a democratic – rather than technocratic – dialogue between citizens and the state about what is acceptable when using algorithmic tools.[418]

Approved Complex Cases could have further requirements. For example, given the rate of technological change, the likelihood of creep towards alternative uses, and the contestable nature of these cases, it would be sensible to require each Complex Case to have a "sunset clause". This would mean that the approval to use the algorithm for the use case would be

---

415 At 12.

416 This provides a tiered approach to various algorithm-related requirements, depending on whether an AIA suggests there is likely to be: (a) little to no impact, which are often reversible and brief; (b) moderate impacts, which are likely reversible and short-term; (c) high impacts, which are likely to be difficult to reverse and are ongoing; or (d) very high impacts, which are likely to be irreversible and perpetual. See Canadian Government, above n 3, Appendix B – Impact Assessment Levels.

417 Gavaghan et al, above n 81, at 74.

418 Notably the Law Society of England and Wales, discussing the use of algorithms in the criminal justice system, suggests that accountability for policy decisions or political design choices which involve difficult trade-offs must not be "outsourced" to third parties. See the Law Society of England and Wales, above n 245, at 4 and 7.

limited to a defined period (e.g., three years) before a new approval would be required from the relevant Minister. The England and Wales Law Society has similarly suggested sunset clauses for uses within the criminal justice system[419] (uses that would typically require Ministerial approval as a Complex Case). This process would require the PSA to review its use of the algorithm and whether it remains consistent with the use case and parameters defined in the AIA. It would also continue the process of political accountability by requiring any new Minister (and by extension, potentially a new Government) to approve the Complex Case.

The above process would also ensure that the use of an algorithm for decision-making is lawful. So, for algorithms that have a non-trivial role in decision-making but which are unlikely to meet the impact threshold above ("**Standard Case**"), use of the algorithm within the defined use case and subject to the mitigations in the ASA would be lawful once: (a) the agency had completed an AIA to confirm this; and (b) the head of the PSA had formally recorded his or her opinion that the use does not meet the statutory threshold for a Complex Case. For a Complex Case, use of an algorithm within the AIA's use case would be lawful once Ministerial approval was given. Any material deviation from the use case would require an agency to go through the same process (effectively, to update the AIA and get any further approvals needed).

The legitimising nature of this process would need to be considered in light of the existing legal protections discussed in this thesis. While the above process would inoculate the PSA from any judicial review on the basis of fettering of discretion or abdication, it should not automatically justify the use of an algorithm where decisions are materially influenced by errors of fact or irrelevant considerations (if the algorithm is used in a way that does not conform with any statutory or common law requirements for the decision). It also would not justify uses *outside* of the use case described in the AIA. Similarly, nothing should prevent a claimant from making a discrimination claim under the HRA. However, where the Minister has approved a Complex Case, and the algorithm is in fact used in accordance with this case, this is likely to provide a strong indication that any discrimination contemplated by the AIA was considered to be "justified" by the Crown.

Likewise, consideration needs to be given to the extent that IPP 8 of the Privacy Act would apply to require the PSA to take reasonable steps to ensure that information relied upon by a decision-maker is accurate. Arguably, what is "reasonable" in terms of accuracy should

---

419 At 5.

also be coloured by any approval given by the Minister for a Complex Case. In relation to an algorithm's outputs (i.e., inferred information), it is unclear that IPP 8 should apply where the Minister has approved a Complex Case which causes harm due to the tuning of false positives or negatives, if the trade-offs are justified in the Minister's mind. On the other hand, IPP 8 *should* apply in those circumstances where the agency is asleep at the wheel, for example where there is a failure to: (a) ensure the accuracy of data inputs into the algorithm; or (b) implement reasonable assurance processes to ensure the algorithm is producing outputs within the range of accuracy, reliability and validity expected from the AIA.

Lastly, the AIA process would need to be subject to any appropriate subject-matter exceptions. For example, while it still makes sense for the intelligence services to have some process for ensuring the validity and reliability of the algorithms they use, this is likely to best sit within the existing frameworks under the Intelligence and Security Act 2017 and potentially the supervision of the Inspector-General of Intelligence and Security.

3 Audit and transparency

The last pillar of the regulatory model would be a process for auditing, and ensuring transparency in, the public sector's use of algorithms.

First, the Algorithms Watchdog would be responsible for holding the AIAs completed by PSAs and lodging these in a register accessible to the public (subject to any redactions necessary under OIA withholding grounds). This register would hold information about: (a) each decision-making algorithm used by an agency; (b) the applicable AIA (including the permitted scope of use and mitigations); and (c) sign off from the Minister or PSA head. The idea of a register of algorithms has been suggested by commentators both in New Zealand and abroad,[420] and would create an important degree of visibility over the areas in which algorithms can impact citizens.

Second, the Algorithms Watchdog should play a role in auditing each agency's use of their algorithms against the AIAs in place, and reporting to Parliament. This would necessitate legal powers to request information and/or visit agencies on site (potentially with a notice period). Audits could cover a range of areas – including software code (if applicable), quality of data, security controls, education and training of decision-makers, the agency's

420 Gavaghan et al, above n 81, at 76; Law Society of England and Wales, above n 245, at 4.

internal assurance processes (e.g., how often and in what manner the PSA checks its algorithms are working appropriately), and "near-misses" and actual incidents of material harm due to algorithmic decision-making. The regularity of review of agencies' use of algorithms would depend on the resourcing of the Algorithms Watchdog. However, ideally the Algorithms Watchdog would report to Parliament annually.

Alongside the Algorithms Watchdog's role, PSAs should be obliged to provide general information about their use of algorithms in decision-making on their public-facing website, in addition to their new obligation to notify individuals of algorithmic decisions (described in chapter three). For example, a PSA should provide the details of the decision-making algorithms it uses, and where the applicable AIAs can be found on the Algorithm Watchdog's register of algorithms. For those algorithms used in relation to a Complex Case – which are likely to be most significant to members of the public – the agency should also provide a plain English explanation of how these algorithms typically make a decision, and explain that individuals have the new rights of notice and human review described in chapter three.

Lastly, the Algorithms Watchdog should issue best practice guidelines for procurement of algorithmic systems. This would help to ensure PSAs meet their obligations under the proposed regulatory model, and that third party providers cannot rely upon legal rights (such as trade secrets) that undermine public transparency. Procurement guidelines should also ensure contracts allow for the auditing and quality assurance processes suggested above. And where applicable, agencies' contracts with suppliers will also need to guarantee access to supplier resources to allow the agency to update the algorithm over time. Last, context-specific requirements may be needed. For example, tools relating to criminal justice might come with the condition that the supplier makes its staff available as expert witnesses in any proceedings regarding how the algorithm works.[421]

E Conclusion

This thesis has traversed three key questions. First, how are algorithms being used in public sector decision-making in New Zealand and abroad, and what are the potential challenges this presents? Second, what legal remedies are available to someone affected by an algorithm used in government decision-making, and where are the gaps in effective

---

421 Babuta et al, above n 22, at 22.

redress? Thirdly, is there a regulatory model that can respond to these gaps so as to provide appropriate protection against harms while optimising the benefits of algorithms?

The discussion in this thesis should make several things clear. First, the use of algorithms in government is only likely to increase, driven primarily by the benefits of service efficiencies and the laudable prospect of overcoming humans' decision-making flaws. Second, the literature indicates a cluster of consistently recurring issues of concern. In particular, using algorithms can create the risk of unfair or biased outcomes, cause privacy intrusions, and can present a challenge to traditional expectations of transparency and natural justice. At the more trivial end, these effects might mean a person is not opted into an automated tax refund. At the other end of the scale, it could mean a person is deprived of his or her personal liberty or deported. These two points – the increasing use of algorithms and their ability to cause material harm – suggests New Zealand's legal framework needs careful attention.

Remedies for those impacted by an algorithm used in decisions may be available in a range of guises, including under the Privacy Act, OIA, HRA, NZBORA, or judicial review. However, this thesis's discussion of these protections reveals their limitations. In particular, these mechanisms are necessarily reactive; they do not *prevent* poor use of algorithms in the first place. Moreover, because of requirements to prove harm and causation, access to a remedy can often depend on the grey line dividing directed decisions and informed decisions. These protections may also struggle to protect against lower-level or cumulative impacts which cause incremental harm. And practical considerations remain important: monetary remedies will not always be substantial or even available, and the cost of bringing a claim can be a significant barrier.

While these protections can be enhanced, the broader solution is a regulatory model that acts as the safety net above the cliff, rather than the ambulance at the bottom. This thesis proposes a regulatory model to mitigate potential harms arising from PSAs' use algorithms, while ensuring transparency over, and political accountability for, these uses. The key to the model is a statutorily independent Algorithms Watchdog to audit, support and report on the public sector's use of algorithms, and a proportional statutory approval process underpinned by the use of AIAs. This model would support the ongoing safe and effective use of algorithms in the New Zealand public sector.

BIBLIOGRAPHY

A Legislation

1 New Zealand

Commerce Act 1986.

Court Matters Act 2018.

District Courts Act 2016.

Environment Act 1986.

Evidence Act 2006.

Financial Markets Authority Act 2011.

Human Rights Act 1993.

Human Rights Amendment Bill 2001 (152-1).

Intelligence and Security Act 2017.

Judicature Amendment Act 1972.

Judicial Review Procedure Act 2016.

Local Government Official Information and Meetings Act 1987.

New Zealand Bill of Rights Act 1993.

Official Information Act 1982.

Ombudsman Act 1975.

Privacy Act 1993.

Privacy Bill 2018 (34-2).

Public Audit Act 2001.

Search and Surveillance Act 2012.

Summary Proceedings Act 1957.

2 Canada

Financial Administration Act (RSC 1985, c, F-11).

3 European Union

Directive 95/46/EC on the protection of individuals with regard to the processing of personal data [1995] OJ L281/31.

Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data [2016] OJ L119/1.

B Cases

1 New Zealand

Air Nelson Ltd v Minister of Transport [2008] NZCA 26, [2008] NZAR 139.

Air New Zealand Ltd v McAlister [2009] NZSC 78, [2010] 1 NZLR 153.

Armfield v Naughton [2014] NZHRRT 48.

Attorney-General v IDEA Services Ltd (In Statutory Management) [2012] NZHC 3229, [2013] 2 NZLR 512.

Attorney-General v Unitec Institute of Technology [2007] 1 NZLR 750 (CA).

Attorney-General v Van Essen [2015] NZCA 22.

Belcher v Chief Executive of the Department of Corrections [2007] 1 NZLR 507 (CA).

Broadcasting Corp of New Zealand v Broadcasting Tribunal [1986] 2 NZLR 620 (CA).

C v Holland [2012] NZHC 2155, [2012] 3 NZLR 672.

Christchurch International Airport Ltd v Christchurch City Council [1997] 1 NZLR 573 (HC).

Combined Beneficiaries Union Inc v Auckland City COGS Committee [2008] NZCA 423, [2009] 2 NZLR 56.

Couch v Attorney-General [2008] NZSC 45, [2008] 3 NZLR 725.

CPAG v Attorney-General [2013] NZCA 420, [2013] 3 NZLR.

CREEDNZ Inc v Governor-General [1981] 1 NZLR 172 (CA).

Criminal Bar Association of New Zealand Incorporated v Attorney-General [2013] NZCA 176.

Curtis v Minister of Defence [2002] 2 NZLR 744 (CA).

Daganayasi v Minister of Immigration [1980] 2 NZLR 130 (CA) at 149.

Director of Human Rights Proceedings v Crampton [2015] NZHRRT 35.

Director of Human Rights Proceedings v Slater [2019] NZHRRT 13.

Federated Farmers of New Zealand Inc v New Zealand Post Ltd [1992] 3 NZBORR 339 (HC).

Forrest v Attorney-General [2012] NZCA 125, [2012] NZAR 798.

Hamilton City Council v Waikato Electricity Authority [1994] 1 NZLR 741 (HC).

Hammond v Credit Union Baywide [2015] HRRT 6.

Harder v Proceedings Commissioner [2000] 3 NZLR 80 (CA).

Hosking v Runting [2005] 1 NZLR 1 (CA).

Kim v Prison Manager Mount Eden Correctional Facility [2012] NZSC 121, [2013] 2 NZLR 589.

Lehmann v Canwest Radioworks Ltd [2006] NZHRRT 35.

Lewis v Wilson & Horton [2000] 3 NZLR 546 (CA).

Mihos v Attorney-General [2008] NZAR 177 (HC).

Ministry of Health v Atkinson [2012] NZCA 184, [2012] 3 NZLR 456.

New Zealand Fishing Industry Association Inc v Minister of Agriculture and Fisheries [1988] 1 NZLR 544 (CA).

Ngaronoa v Attorney-General; Taylor v Attorney-General [2017] NZCA 351, [2017] 3 NZLR 643.

NRHA v Human Rights Commission [1998] 2 NZLR 218 (HC).

Pharmaceutical Management Agency Ltd v Roussel Uclaf Australia Pty Ltd [1998] NZAR 58 (CA).

Poamanga v State Services Commission [1985] 2 NZLR 385 (CA).

Quake Outcasts v Minister for Canterbury Earthquake Recovery [2015] NZSC 27, [2016] 1 NZLR 1.

Quilter v Attorney-General [1998] 1 NZLR 523 (CA).

Hamed v R [2011] NZSC 101, [2012] 2 NZLR 303.

R v Alsford [2017] NZSC 42, [2017] 1 NZLR 710.

R v Hansen [2007] NZSC 7, [2007] 3 NZLR 1.

Re Vixen Digital Limited [2003] NZAR 418 (HC).

Ririnui v Landcorp Farming [2016] NZSC 62.

Simpson v Attorney General [1994] 3 NZLR 667 (CA).

Small v Attorney-General (2000) 6 HRNZ 218 (HC).

Smith v Air New Zealand Ltd [2011] NZCA 20, [2011] 2 NZLR 171.

Spencer v Ministry of Health [2016] NZHC 1650, [2016] 3 NZLR 513.

Spencer v Ministry of Health [2017] NZHC 291.

Tapiki and Eru v New Zealand Parole Board [2019] NZHRRT 5.

Tan v New Zealand Police [2016] NZHRRT 32.

Tannadyce Investments Ltd v Commissioner of Inland Revenue [2011] NZSC 158, [2012] 2 NZLR 15.

Taylor v Attorney-General [2016] NZHC 355, [2016] 3 NZLR 111.

Taylor v Chief Executive of the Department of Corrections [2015] NZCA 477, [2015] NZAR 1648.

Television New Zealand Ltd v West [2011] 3 NZLR 825 (HC).

Waitakere City Council v Lovelock [1997] 2 NZLR 385 (CA).

Westhaven Shellfish Ltd v Chief Executive of Ministry of Fisheries [2002] 2 NZLR 158 (CA).

Wilson v NZ Customs Service (1999) 5 HRNZ 134 (HC).

XY v Attorney General [2016] NZHC 1196, [2016] NZAR 875.

Ye v Minister of Immigration [2009] 2 NZLR 596 (CA).

2 England and Wales

British Oxygen Co Ltd v Minister of Technology [1971] AC 610 (HL).

Patmalniece v Secretary of State for Work and Pensions [2011] UKSC 11, [2011] 1 WLR.

R (Bridges) v CCSWP and SSHD [2019] EWHC 2341 (Admin).

R v Panel on Take-overs and Mergers, ex parte Guinness Plc [1990] 1 QB 146, [1989] 1 All ER 509 (CA).

3 Canada

British Columbia (Public Service Employee Relations Commission) v BCGEU [1999] 3 SCR 3.

Eldridge v British Columbia (Attorney-General) [1997] 3 SCR 624.

Ewert v Canada 2018 SCC 30.

Slaight Communications Inc v Davidson [1989] 1 SCR 10.

Wynberg v Ontario (2006) 269 DLR (4th) 435 (ONCA).

4 USA

State v Loomis 881 NW 2d 749 (Wis 2016).

KW v Armstrong No. 14-35296 (9th Cir 2015).

C Books

Etham Alpaydin *Introduction to Machine Learning* (MIT Press, Cambridge, 2014).

Andrew Butler and Petra Butler *New Zealand Bill of Rights Act: A Commentary* (2nd ed, LexisNexis, Wellington, 2015).

Virginia Eubanks *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor* (St Martin's Press, New York, 2017).

Hannah Fry *Hello World: How to Be Human in the Age of the Machine* (Transworld Publishers, London, 2018).

Daniel Kahneman *Thinking, Fast and Slow* (Penguin Books, London, 2012).

Michael Lewis *Moneyball: The Art of Winning an Unfair Game* (W W Norton & Company, New York, 2004).

Paul E Meehl *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (University of Minnesota Press, Minneapolis, 1954).

Cathy O'Neil *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Penguin Books, New York, 2016).

Frank Pasquale *The Black Box Society: The Secret Algorithms that Control Money and Information* (Harvard University Press, Cambridge, 2015).

Matthew Smith *New Zealand Judicial Review Handbook* (2nd ed, Thomson Reuters New Zealand Ltd, Wellington, 2016).

Graham Taylor *Judicial Review: A New Zealand Perspective* (4th ed, LexisNexis NZ Limited, Wellington, 2018).

Richard H Thaler and Cass R Sunstein *Nudge: Improving Decisions about Health, Wealth and Happiness* (Yale University Press, 2008).

Joseph Turow *The Daily You: How the New Advertising Industry is Defining Your Identity and Your Worth* (Yale University Press, New Haven, 2012).

D Chapters in edited books

Allesandro Acquisti and Jens Grossklags "What can Behavioural Economics Teach Us about Privacy" in Allesandro Acquisti and others (eds) *Digital Privacy: Theory, Technologies, and Practices* (CRC Press, New York, 2007) 363.

Alvaro M Bedoya "Algorithmic Discrimination vs Privacy Law" in Evan Selinger, Jules Polonetsky and Omer Tene (eds) *The Cambridge Handbook of Consumer Privacy* (Cambridge University Press, Cambridge, 2018) 232.

Andrew Le Sueur "Robot Government: Automated Decision-making and its Implications for Parliament" in Alexander Horne and Andrew Le Sueur (eds) *Parliament: Legislation and Accountability* (Hart Publishing, Oxford, 2016) 183.

Mark MacCarthy "In Defense of Big Data" in Evan Selinger, Jules Polonetsky and Omer Tene (eds) *The Cambridge Handbook of Consumer Privacy* (Cambridge University Press, Cambridge, 2018) 47.

A Naryanan, J Huey and E Felten "A Precautionary Approach to Big Data Privacy" in S Gutwirth, R Leenes and P De Hert (eds) *Data Protection on the Move: Current Developments in ICT and Privacy/Data Protection* (Springer, Berlin, 2016) 357.

Christopher Wolf "Envisioning Privacy in the World of Big Data" in Marc Rotenberg, Jeramie Scott and Julia Horwitz (eds) *Privacy in the Modern Age: The Search for Solutions* (The New Press, 2015) 204.

E Looseleaf texts

Paul Roth *Privacy Law and Practice* (online looseleaf ed, LexisNexis NZ Limited).

F Journal articles

Philip Adler, Casey Falk, Sorelle A Friedler, Tionney Nix, Gabriel Rybeck, Carols Scheidegger, Brandon Smith and Suresh Venkatasubramanian "Auditing Black-box Models for Indirect Influence" (2018) 54 Knowl Inf Syst 95.

Jack M Balkin "2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data" (2017) 78 Ohio St LJ 1217.

Solon Barocas and Helen Nissenbaum "Big Data's End Run Around Procedural Privacy Protections" (2014) 57 Communications of the ACM 31.

Solon Barocas and Andrew D Selbst "Big Data's Disparate Impact" (2016) 104 Calif L Rev 671.

Gary D Bass "Big Data and Government Accountability: An Agenda for the Future" (2013) 11 ISLJP 13.

Samuel Beswick "Perlustration in the Pathless Woods: *Hamed v R*" (2011) 17 Auck U L Rev 291.

Jenna Burrel "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms" (January – June 2016) Big Data & Society 1.

Angele Christin "Algorithms in Practice: Comparing Web Journalism and Criminal Justice" (July – December 2017) Big Data & Society 1.

Danielle Keats Citron and Frank A Pasquale "The Scored Society: Due Process for Automated Predictions" (2014) 89 Wash L Rev 1.

Luiz Costa "Privacy and the Precautionary Principle" (2012) 28 Comp Law & Sec Rev 14.

Stephanie Cuccaro-Alamin, Regan Foust, Rhema Vaithianathan, Emily Putnam-Hornstein "Risk Assessment and Decision-Making in Child Protective Services: Predictive Risk Modelling in Context" (2017) 79 Children and Youth Services Review 291.

Rashmi Dayalu, Elizabeth T Cafiero-Fonseca, Victoria Y Fan, Heather Schofield and David E Bloom "Priority Setting in Health: Development and Application of a Multi-criteria Algorithm for the Population of New Zealand's Waikato Region" (2018) 16 Cost Eff Resour Alloc 35.

Nicholas Diakopoulous "Accountability in Algorithmic Decision Making" (2016) 59 Communications of the ACM 56.

Lilian Edwards & Michael Veale "Slave to the Algorithm? Why a 'Right to an Explanation' is Probably not the Remedy you are Looking for" (2017) 16 Duke Law & Tech Rev 18.

Stefanía Ægisdóttir, Michael J White, Paul M Spengler, Alan S Maugherman, Linda A Anderson, Robert S Cook, Cassandra N Nichols, Georgios K Lampropoulos, Blain S Walker, Genna Cohen and Jeffrey D Rush "The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Acculmulated Research on Clinical Versus Statistical Prediction" (2006) 34 The Counselling Pyschologist 341.

Asher Gabriel Emanuel "To Whom Will Ye Liken Me, and Make Me Equal? Reformulating the Role of the Comparator in the Identification of Discrimination" (2014) 45 VUWLR 1.

Katrine Evans "Show Me the Money: Remedies Under the Privacy Act" (2005) 36 VUWLR 475.

Joshua AT Fairfield and Christoph Engel "Privacy as a Public Good" (2015) 65 Duke LJ 385.

Andrew G Ferguson "Big Data and Predictive Reasonable Suspicion" (2015) 163 U Pa L Rev 327.

Anthony W Flores, Kristin Bechtel and Christopher T Lowenkamp "False Positives, False Negatives, and False Analyses: A Rejoinder to 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks.'" (2016)

80 Federal Probation 38.

Katherine Freeman “Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v Loomis” (2016) 18 NCJL & Tech 75.

Oscar H Gandy, Jr “Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems,” (2010) 12 Ethics and Information Technology 29.

Claudia Geiringer “Sources of Resistance to Proportionality Review of Administrative Power under the New Zealand Bill of Rights Act” (2013) 11 NZJPIL 123.

Lewis R Goldberg “Man versus Model of Man: A Rationale, Plus Some Evidence, for a Method of Improving on Clinical Inferences” (1970) 73 Psychological Bulletin 422.

Lewis R Goldberg “Simple Models or Simple Processes? Some Research on Clinical Judgments” (1968) 23 American Psychologist 483.

Maria Eduarda Goncalves “The Risk-Based Approach under the New EU Data Protection Regulation: a Critical Perspective” (2019) Journal of Risk Research 1.

Riccardo Guidotti, Anna Monreale, Salvatorre Ruggieri, Franco Turini, Fosca Giannotti and Dino Pedreschi “A Survey of Methods for Explaining Black Box Models” (2018) 51 ACM Computing Surveys, Article 93.

H Brian Holland “Privacy Paradox 2.0” (2010) 19 Widener LJ 893.

Philip N Howard, Samuel Wooley and Ryan Calo “Algorithms, Bots and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration” (2018) 15 Journal of Information Technology & Politics 81.

Elizabeth E Joh “The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing” (2016) 10 Harv L & Pol’y Rev 15.

Emily Keddell “The Ethics of Predictive Risk Modelling in the Aotearoa/New Zealand Child Welfare Context: Child Abuse Prevention or Neo-liberal Tool?” (2014) 35 Critical

Social Policy 69.

Dean R Knight "Mapping the Rainbow of Review: Recognising Variable Intensity" (2010) NZ L Rev 393.

David Kriebel, Joel Tickner, Paul Epstein, John Lemons, Richard Levins, Edward L Loechler, Margaret Quinn, Ruthann Rudel, Ted Schettler and Michael Stoto "The Precautionary Principle in Environmental Science" (2002) 109 Environmental Health Perspectives 871.

Joy Liddicoat, Colin Gavaghan, Alistair Knott, James Maclaurin and John Zerilli "The Use of Algorithms in the New Zealand Public Sector" (2019) NZLJ 26.

Min Kyung Lee "Understanding Perception of Algorithmic Decisions: Fairness, Trust and Emotion in Response to Algorithmic Management" (January – June 2018) Big Data & Society 1.

Bruno Lepri, Nuria Oliver, Emmanuel Letouze, Alex Pentland and Patrick Vinck "Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges" (2017) 31 Philos Technol 611.

Alessandro Mantelero "Regulating Big Data: The Guidelines of the Council of Europe in the Context of the European Data Protection Framework" (2017) 33 Comp Law & Sec Rev 584.

A Jay McClurg "Bringing Privacy Law out of the Closet: A Tort Theory of Liability for Intrusions in Public Places" (1995) 73 North Carolina L Rev 989.

Danica McGovern "Ewert v Canada (2018) SCC 30" [2019] NZLJ 131.

Brent Mittelstadt "Auditing for Transparency in Content Personalization Systems" (2016) 10 International Journal of Communication 4991.

Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter and Luciano Floridi "The Ethics of Algorithms: Mapping the Debate" (July – December 2016) Big Data & Society 1.

N A Moreham “Privacy in Public Places” (2006) 65 Cambridge Law Journal 606.

S C Olhede and P J Wolfe “The Growing Ubiquity of Algorithms in Society: Implications, Impacts and Innovations” (2018) 376: 20170364 Phil Trans R Soc A 1.

Linda Onnasch, Christopher D Wickens, Huiyang Li and Dietrich Manzey “Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis” (2014) 56 Hum Factors: J Human Fact Ergon Soc 476.

Marion Oswald “Algorithm-assisted Decision-making in the Public Sector: Framing the Issues using Administrative Law Rules Governing Discretionary Power” (2018) 376: 20170359 Phil Trans R Soc A 1.

Marion Oswald, Jamie Grace, Sheena Urwin and Geoffrey C Barnes “Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and ‘Experimental’ Proportionality” (2018) 27 Information and Communications Technology Law 223.

Raja Parasuraman and Dietrich H Manzey “Complacency and Bias in Human Use of Automation: An Attentional Integration” (2010) 52 Hum Factors: J Human Fact Ergon Soc 381.

Frank Pasquale “Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society” (2017) 78 Ohio St LJ 1243.

Kayvan Pazouki, Neil Forbes, Rosemary A Norman and Michael D Woodward “Investigation on the Impact of Human-automation Interaction in Maritime Operations” (2018) 153 Ocean Engineering 297.

Steven Price “The Official Information Act: Does it Work?” (2016) NZLJ 276.

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

Neil Richards “The Dangers of Surveillance” (2013) 126 Harv L Rev 1935.

Andrew D Selbst “Disparate Impact in Big Data Policing” (2017) 52 Ga L Rev 109.

Latanya Sweeney “Discrimination in online ad delivery” (2013) 11 ACM Queue 10.

Sonja B Starr "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination" (2014) 66 Stan L Rev 803.

Sonja B Starr "The Odds of Justice: Actuarial Risk Prediction and the Criminal Justice System" (2016) 29 Chance 49.

Evan Selinger and Kyle White “Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture” (2011) 10 Sociology Compass 923.

Michael Taggart “Proportionality, Deference, Wednesbury” (2008) NZ L Rev 423.

Mahoney Turnbull “Navigating New Zealand’s Digital Future: Coding our Way to Privacy in the Age of Analytics” (2015) 3 NZLSJ 420.

Andrew Tutt “An FDA for Algorithms” (2017) 69 Admin L Rev 83.

Amos Tversky and Daniel Kahneman “Judgement under Uncertainty: Heuristics and Biases” (1974) 185 Science 1124.

Rhema Vaithianathan “Children in the Public Benefit System at Risk Maltreatment: Identification via Predictive Modelling” (2013) 45 Am J Prev Med 354.

Karen Yeung “Algorithmic Regulation: A Critical Interrogation” (2018) 12 Regulation & Governance 505.

Karen Yeung “’Hypernduge’: Big Data as a Mode of Regulation by Design” (2017) 20 Information, Communication & Society.

John Zerilli, Alistair Knott, James Maclaurin and Colin Gavaghan “Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?” (September 2018) *Philosophy & Technology* 1.

G Conference and seminar papers

Peter Barnett "Remedies for Discrimination by Government under Part 1A of the Human Rights Act 1990" (paper presented to the New Zealand Law "Society Using Human Rights Law in Litigation" Intensive Conference, June 2014) 135.

Francis Cooke "Judicial Review" (New Zealand Law Society seminar, May 2012).

Dr Rodney Harrison QC "Remedies for Breach of the New Zealand Bill of Rights Act 1990: The New Zealand Experience – Recognising Rights While Withholding Meaningful Remedies" (paper presented to the New Zealand Law "Society Using Human Rights Law in Litigation" Intensive Conference, June 2014) 107.

Grant Illingworth QC "Discretion, Legality and the Bill of Rights" (paper presented to the New Zealand Law "Society Using Human Rights Law in Litigation" Intensive Conference, June 2014) 1.

Christian Sandvig, Kevin Hamilton, Karrie Karahalios and Cedric Langbort "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms" (paper presented to the Data and Discrimination: Converting Critical Concerns into Productive Inquiry preconference of the 64th Annual Meeting of the International Communication Association, Seattle 22 May 2014).

H Unpublished papers

Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayan Irshad and Andrew H Beck "Deep Learning for identifying metastatic breast cancer" Cornel University Library (18 June 2016).

Moritz Hardt, Eric Price and Nathan Srebro "Equality of Opportunity in Supervised Learning" (7 October 2016).

Jon Kleinberg, Sendhil Mullainathan and Manish Raghavan "Inherent Trade-Offs in the Fair Determination of Risk Scores" (17 November 2016).

Teresa Scantamburlo, Andrew Charlesworth and Nello Cristianini "Machine Decisions and

Human Consequences" (2018) (draft chapter for the forthcoming book K Yeung and M Lodge (eds) *Algorithmic Regulation*).

I Reports and guidance

Alexander Babuta, Marion Oswald and Christine Rinik *Machine Learning Algorithms and Police Decision-Making Legal, Ethical and Regulatory Challenges: Whitehall Report 3-18* (September 2018).

AI Forum New Zealand *Artificial Intelligence: Shaping a Future New Zealand* (March 2018).

AI Now *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability* (April 2018).

Anton Blank, Fiona Cram, Tim Dare, Irene De Haan, Barry Smith and Rhema Vaithianathan *Ethical Issues for Maori in Predictive Risk Modelling to Identify New-Born Children who are at Risk of Future Maltreatment* (January 2015).

Article 29 Data Protection Working Party *Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing is "Likely to Result in a High Risk" for the purpose of Regulation 2016/679* (4 October 2017, 17/EN WP 248 rev.01).

Australian Government *Automated Assistance in Administrative Decision-Making* (February 2007).

Australian Government *Artificial Intelligence: Australia's Ethics Framework – A Discussion Paper* (Canberra, April 2019).

Canadian Government *Directive on Automated Decision-Making* (April 2019).

Tim Dare *Predictive Risk Modelling and Child Maltreatment: An Ethical Review* (25 September 2013).

Data Futures Partnership *A Path to Social Licence: Guidelines for Trusted Data Use* (August 2017).

Data Futures Partnership and Toi Aria (Massey University) *Our Data, Our Way: What New Zealand People Expect from Guidelines for Data Use and Sharing – Findings from Public Engagement* (March 2017).

Federal Trade Commission *Big Data: A Tool for Inclusion or Exclusion – Understanding the Issues* (January 2016).

Freedom House *Freedom on the Net: Manipulating Social Media to Undermine Democracy* (November 2017).

Colin Gavaghan, Alistair Knott, James Maclaurin, John Zerilli and Joy Liddicoat *Government Use of Artificial Intelligence in New Zealand: Final Report on Phase 1 of the New Zealand Law Foundation's Artificial Intelligence and Law in New Zealand Project* (Wellington, 2019).

The White House *Big Data: Seizing Opportunities, Preserving Values* (Washington, May 2014).

The White House *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (Washington, May 2016).

Georgetown Law Centre on Privacy & Technology *The Perpetual Line Up: Unregulated Police Face Recognition in America* (Washington, October 2016).

The Law Society of England and Wales *Algorithms in the Criminal Justice System* (June 2019).

House of Commons Science and Technology Committee *Algorithms in Decision-making: Fourth Report of Session 2017-19* (15 May 2018).

House of Commons Science and Technology Committee *The Big Data Dilemma: Fourth Report of Session 2015-16* (10 February 2016).

House of Lords Select Committee on Artificial Intelligence *AI in the UK: Ready, Willing and Able?: Report of Session 2017 – 19* (16 April 2018).

Law Commission *Review of the Privacy Act 1993: Review of the Law of Privacy Stage 4* (NZLC R123, 2011).

Ministry of Social Development *The Feasibility of Using Predictive Risk Modelling to Identify New-born Children who are High Priority for Preventive Services* (Wellington, 2 February 2014).

Ministry of Social Development *The White Paper for Vulnerable Children: Volume I* (October 2012).

Ministry of Social Development *The White Paper for Vulnerable Children: Volume II* (October 2012).

New Zealand Human Rights Commission *Privacy, Data and Technology: Human Rights Challenges in the Digital Age* (Auckland, May 2018).

New Zealand Treasury *Environmental Risk Management in New Zealand: Is there a Scope to Apply a More Generic Framework? – Policy Perspectives Paper* (Wellington, July 2006).

OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980).

Office of the Australian Information Commissioner *Guide to Data Analytics and the Australian Privacy Principles* (March 2018).

Office of the Prime Minister's Chief Science Advisor *Using Evidence to Inform Social Policy: the Role of Citizen-based Analytics* (Auckland, 19 June 2017).

Office of the Privacy Commissioner *Inquiry into the Ministry of Social Development's Collection of Individual Client-Level Data from NGOs* (4 April 2017).

Office of the Privacy Commissioner *Privacy Impact Assessment Toolkit* (July 2015).

Office of the Privacy Commissioner and Statistics New Zealand *the Principles for Safe and Effective Use of Data and Analytics* (16 May 2018).

Osonde Osoba and Willam Wesler *An Intelligence in our Image* (Rand Corporation, Santa Monica, 2017).

Emily Putnam-Hornstein and Tim Dare *Vulnerable Children: Can Administrative Data be Used to Identify Children at Risk of Adverse Outcomes* (September, 2012).

Lee Rainie and Janna Anderson *Code Dependent: The Pros and Cons of the Algorithmic Age* (Pew Research Centre, 8 February 2017).

The Royal Society *Machine Learning: The Power and Promise of Computers that Learn by Example* (London, April 2017).

Social Investment Agency *What You Told Us: Findings of the 'Your Voice, Your Data, Your Say' Engagement on Social Wellbeing and the Protection and Use of Data* (November 2018).

Statistics New Zealand and Department of Internal Affairs *Algorithm Assessment Report* (Wellington, 2018).

Waitangi Tribunal *The Offender Assessment Policies Report* (Wai 1024, 2005).

The Workshop *Digital Threats to Democracy* (May 2019).

J Newspaper and magazine articles

Julia Angwin and Jeff Larson "Bias in Criminal Risk Score is Mathematically Inevitable Researchers Say" *ProPublica* (30 December 2016).

Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks" *ProPublica* (23 May 2016).

Gill Bonnett "Immigration NZ Using Data System to Predict Likely Troublemakers" *Radio New Zealand* (5 April 2018).

Robert Booth "Benefits System Automation Could Plunge Claimants Deeper into Poverty"

The Guardian (14 October 2019).

Sam Corbett-Davies, Emma Pierson, Avi Feller and Sharad Goel "A Computer Program Used for Bail and Sentencing Decisions was Labelled Biased against Blacks. It's Actually Not That Clear." *The Washington Post* (17 October 2016).

Simon Elvery "How Algorithms Make Important Government Decisions – and How That Affects You" *ABC News* (21 July 2017).

Cyrus Farivar "New Bill Aims to Stamp out Bias in Algorithms Used by Companies" NBC News (11 April 2019).

Alex Hern "How Social Media Filter Bubbles and Algorithms Influence the Election" *The Guardian* (22 May 2017).

Zheping Huan "All Chinese Citizens Now Have a Score Based on How Well We Live and Mine Sucks" *Quartz* (10 October 2015).

Ellora Thadaney Israni "When an Algorithm Helps Send You to Prison" *The New York Times* (October 26, 2017).

Kirsty Johnston "Privacy and Profiling Fears over Secret ACC Software" *New Zealand Herald* (15 September, 2017).

Daniel Kahneman, Andrew M Rosenfield, Linnea Gandhi and Tom Blaser "Noise: How to Overcome the High, Hidden Costs of Inconsistent Decision-Making" *Harvard Business Review* (October 2016).

Stacey Kirk "Children 'Not Lab-rats' – Anne Tolley Intervenes in Child Abuse Experiment" *Stuff* (30 July 2015).

Nicole Kobie "The Complicated Truth About China's Social Credit System" *Wired* (21 January 2019).

Jeff Larson, Surya Mattu, Lauren Kirchnet and Julia Angwin "How We Analysed the COPMAS Recidivism Algorithm" *Propublica* (23 May 2016).

Sam Levin “New AI Can Guess Whether You’re Gay or Straight from a Photograph” *The Guardian* (8 September 2017).

Emma Martinho-Truswell “How AI Could Help the Public Sector” *Harvard Business Review* (29 January 2018).

Asha McLean “NZ Immigration Rejects 'Racial Profiling' Claims in Visa Data-modelling Project” *ZDNet* (6 April 2018).

Joel McManus “Why a Pastor who Abused Children Served Half as Much Prison Time as a Low-level Cannabis Dealer” *Stuff* (13 August 2019).

Tze Ming Mok “Crap Models and Laughable Claims: Immigration NZ’s Spreadsheet Fiasco” *The Spinoff* (10 April 2018).

“Mnangawa Invests in Super Spyware for Citizens” *Zambezi Post* (29 June 2019).

Michael Neilson “Artificial Intelligence Assists with Heart Attack Diagnosis in NZ-led Research” *New Zealand Herald* (11 September 2019).

Frank Pasquale “Secret Algorithms Threaten the Rule of Law” *MIT Technology Review* (1 June 2017).

“Privacy Commissioner Would be ‘Very Worried’ if Auckland Transport Introduces Facial Recognition Technology in Cameras” *1News* (13 August 2019).

“Robo-debt Class Action Could Deliver Justice for Tens of Thousands of Australians Instead of Mere Hundreds” *The Conversation* (17 September 2019).

Sigal Samuel “A New Study Finds a Potential Risk With Self-Driving Cars: Failure to Detect Dark-skinned Pedestrians” *Vox* (6 March 2019).

Frith Tweedie “CIO Upfront: She’ll Be Right? Government Review of Algorithms Shows Need for Caution” *CIO New Zealand* (3 December 2018).

Ali Winston “Palantir has Secretly Been using New Orleans to Test its Predictive Policing Technology” *The Verge* (27 February 2018).

K Websites

Department of Corrections "Risk of Reconviction" <https://www.corrections.govt.nz/resources/research_and_statistics/risk-of-reconviction.html>.

New York City Automated Decision Systems Task Force "About" <https://www1.nyc.gov/site/adstaskforce/about/about-ads.page>.

New Zealand Government "Government Chief Digital Officer" (May 2019) <https://www.digital.govt.nz/digital-government/leadership-and-governance/government-chief-digital-officer-gcdo>.

State Services Commission "What is the 'Public Sector'" (April 2018) <http://www.ssc.govt.nz/resources/what-is-the-public-sector/>.

United Kingdom Government "Centre of Data Ethics and Innovation" <https://www.gov.uk/government/groups/centre-for-data-ethics-and-innovation-cdei>.

L Blogs

Jay Stanley "Pitfalls of Artificial Intelligence Decisionmaking Highlighted in Idaho ACLU Case" (2 June 2017) American Civil Liberties Union <https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case>.

M Letters and submissions

John Edwards "Privacy Commissioner's Submission on the Privacy Bill to the Justice and Electoral Select Committee" (31 May 2018).

Letter from Rodger Haines QC (Chairperson of the Human Rights Review Tribunal) to Andrew Little (Minister of Justice) regarding the Human Rights Review Tribunal (3 November 2017).