Open Rights or Secret Risk Assessments? New Challenges for Public Law in an Age of Artificial Intelligence and the Law

Keynote Speech

Patrick Keyzer[1]

[1] La Trobe University, Bundoora, VIC 3086, Australia
p.keyzer@latrobe.edu.au

1 The Cautionary Tale of "Robo-Debt"

Centrelink, Australia's welfare agency, recently contacted the Commonwealth Scientific and Industrial Research Organisation, Australia's leading public scientific agency, and asked its "Data61" team to review its electronic data-matching system with the Australian Taxation Office, the "Online Compliance Intervention". The purpose of the Online Compliance Intervention was to data-match tax records and welfare payment records, work out whether there were any discrepancies, and then correct them (Commonwealth Ombudsman, 2017). It makes sense for governments to ensure that welfare benefits are paid to people who are genuinely in need, and to ensure that benefits are not paid to people who have paid employment. Properly set up and operated, such a system could help ensure that resources are most effectively deployed to prevent poverty and minimise welfare fraud.

However this is not how things worked in practice, in a great many cases. Instead, from about midway through last year and well into this year, Centrelink's data-matching system was conflating annual earnings figures from the Tax Office with fortnightly income profiles used by Centrelink to assess welfare payment needs. It transpires that the "Matching Techniques" Protocol deployed an algorithm to calculate totals for a number of financial years, but since Centrelink doesn't return detailed fortnightly data about a given *year's* earnings, the algorithm simply averaged the amounts over the relevant periods. This approach was prone to errors, particularly in circumstances where a beneficiary might have made transitions in and out of work (increasingly common in a casualised labour market).

Regrettably, the errors produced by this defective approach were compounded by the administrative approach taken to the recoupment of the debts. Letters were sent out automatically, without any internal, human review of the relevant calculations. These identified debts of hundreds or even thousands of dollars and then required the recipients to contact Centrelink within 21 days to explain the discrepancy, or pay the debt under penalty of enforcement. The letters did not include contact telephone numbers for the compliance team, so people seeking assistance contacted Centrelink through a general call line that resulted in long waiting times. It transpires that the

call line was staffed by people who had little knowledge of the Online Compliance Intervention because they had not been trained.

The characterisation of these letters has been a bone of contention. They were described by the departmental secretary in a subsequent Senate Inquiry as "clarification letters" – as providing *opportunities* for welfare beneficiaries to "clarify" the discrepancy between the identified debt and the beneficiary's personal records. However to ensure that this "clarification" was provided within a particular time period, the Government authorised private debt collectors to follow-up the letters with telephone calls in return for a 10% bonus payment. The "clarification" was thus outsourced by the government department to the "customer", and to be conducted under the shadow of a penalty.

After many complaints, the Commonwealth Ombudsman initiated an investigation. The Ombudsman's report concluded that the enforcement regime imposed unreasonable burdens on welfare beneficiaries and Centrelink staff. The Ombudsman said that while it was reasonable for Centrelink to ask beneficiaries for assistance in explaining discrepancies in its records, that the 21-day timeframe was unreasonable, and that the success of the Initiative depended on its usability. Usability in turn depended on the accuracy and completeness of departmental information, which was questionable in many cases. (Peter Hanks QC has questioned this, noting that the Ombudsman did not address the question whether the Social Security Act can create a debt presumptively, and whether the Department of Human Services could shift the fact-finding task to the individual: Hanks, 2017). The Ombudsman also found that the requirement that people keep records over six or seven years was unreasonable, in part because beneficiaries had not been forewarned of this requirement. The Report also outlined problems related to planning and implementation, lack of consultation, and a failure to plan for or properly mitigate risks. In addition, the Ombudsman found that Centrelink's assistance and customer support was defective (it has been reported that there were 42 million unanswered phone calls in a month) and that staff had not been adequately trained to support customers and to deal with complaints.

The Robo-Debt Controversy could be dismissed as a classic case of "garbage in, garbage out" – the "algorithm" was flawed, and if it was fixed, then, according to the Ombudsman, it was otherwise fair for Centrelink to request "clarification". Perhaps the reference to Data 61 means that the Government will take expert advice to ensure that the design of such systems will be improved in the future. But the Robo-Debt Controversy is a cautionary tale. With the increasing use of artificial intelligence systems in public administration, where is the place for human oversight and procedural justice? What steps should we take to protect the most vulnerable people in our society from the public service machines? This paper raises some questions. Answering them will take more work.

***

The deployment of artificial intelligence systems makes a lot of sense in large-scale, routine work that requires no or only minimal and manageable discretion (Perry and Smith, 2014). For example, a program developed by the United States Depart-

ment of Veterans Affairs manages disability claims and has completely replaced human public servants by requiring applicants to fill in a detailed questionnaire that is then processed by software. Properly designed and implemented, these systems can speed up the operation of administrative justice and enhance transparency. Questions arise, of course, when artificial intelligence initiatives disrupt—legally-sanctioned bureaucratic authority. Notwithstanding the risks, it is likely that governments will continue to seek out new ways to automate such systems, in order to save revenue, particularly in environments of austerity (Perry and Smith, 2014) (let alone artificially-produced scarcity).

In a recent lecture series at Penn State, Justice Cuellar of the California Supreme Court, a former Professor of Law and Information Technology at Stanford Law School, has identified a number of potential side effects from automated public decision-making. Cybersecurity risks are an obvious example. However the impact of automation on *dialogue* is by far the most important:

> Implicit in democratic governance is an aspiration for dialogue and exchange of reasons that are capable of being understood, accepted, or rejected by policymakers, representatives of organized interests, and members of the public.
>
> Except when computerized decisions can rely on relatively straightforward, rule-like structures, difficulties will arise in supplying explanations of how decisions were made that could be sufficiently understood by policymakers and the public.

Cuellar (2016) also remarked:

> This is not to say that the status quo is any deliberative panacea. On the contrary, it is easy to criticize the current administrative state for its lack of opportunities to allow the public to participate in decisions. Yet the growing reliance on automated computer programs to make sensitive decisions in the administrative state will only complicate what little deliberation does occur.

Do the principles of Australian law provide adequate normative resources for dealing with the challenges ahead? This is a question that could inspire many academic papers. In this paper I will focus on just one, but an important one: What happens when we don't know why the machine has made the decision it has made?

2 Risk Assessment on Secret Grounds

In 'The Minority Report' (1956), the science fiction writer Philip K Dick famously invented 'Precrime', a government agency (later popularized by Tom Cruise in a very ordinary movie) which enabled the surveillance and apprehension of people who would commit murders in the future. Today, suspected terrorists can be detained without charge on suspicion of future harm and sex offenders can be sent to prison on the basis of a risk assessment in circumstances where they have committed no new

crime (McSherry and Keyzer, 2009). The United Nations Human Rights Committee has said in several decisions that imprisoning a person on the basis of a risk assessment in the absence of a fresh crime and criminal trial is arbitrary detention and incompatible with the International Covenant on Civil and Political Rights (Keyzer, 2011) but this continues to be done, as Australia does not honour its international human rights obligations (O'Donovan and Keyzer, 2014). In the absence of human rights norms, Australia is left with weak constitutional protections and the common law. Can these principles ensure procedural justice in pre-crime scenarios?

The use of secret algorithms in risk assessments is a new, worrying development in the criminal justice system. Recently, a Wisconsin trial court sentenced a man named Eric Loomis to six years' imprisonment for participating in a drive-by shooting (Liptak, 2017). In sentencing, the trial court considered a report derived from a software product called Compas, produced by a company called Northpointe Inc. Compas uses an algorithm to weigh a number of risk factors and produce an actuarial risk assessment of a person accused of a crime. At trial, the prosecutor submitted a Compas report about Loomis that found him to be a high risk of violence, a high risk of recidivism, and a high pretrial risk. Loomis sought details about the software algorithm so that he could challenge its conclusions. Northpointe declined to release any details about how its Compas algorithm calculates risk on the basis that it is proprietary, and commercial-in-confidence. Loomis appealed the ruling of the trial judge.

The process of actuarial risk assessment is well explained by Brad Johnson in his paper "Prophecy With Numbers" (2006), in terms worth setting out at some length:

> Psychiatrists and psychologists have employed a number of methods for determining risk with respect to human behaviour, which include … clinical assessment, actuarial risk assessment and actuarially informed clinical assessment which combines elements from each. The difference between clinical and actuarial assessment is reflected in the type of data relied on in order to determine the level of risk—clinical assessment relying primarily on data about the person being assessed and actuarial assessment relying on data from a population of individuals who share a number of attributes in common with the person being assessed, thus allowing statistical comparative judgements. …
>
> Actuarial risk assessment departs from clinical assessment methods by examining populations of released offenders in order to identify attributes that are associated with an increased risk of recidivism. The data with respect to recidivism rates collected from multiple sample populations of released offenders can be used to make some simple inferences. The relative frequency of recidivism for a particular sample may be used to make a probability statement about the chance of an individual, who shares the attributes that define the population, committing a future offence. Alternatively the relative frequencies for various populations may be compared to determine which samples display a higher level recidivism, which in turn is believed to indicate a greater risk of recidivism. The process of establishing relative frequencies with respect to recidivism begins by examining an initial population of released offenders for a specific period of time which yields relative frequencies for those who reoffend and those who do not. The sample population being investigated also allows researchers to look for attributes that are associated with recidivism. The initial population can

> then be analysed by specifying further attributes that break the population down into more clearly defined demographic groups in the hope of identifying greater recidivism rates for specific populations.

Elsewhere, Bernadette McSherry and I have written that the use of risk assessment scales may be justifiable for the purpose of treatment in clinical environments (2009), but problems with these scales are amplified by their use in legal forums, where they can be 'prone to manipulation and misinterpretation' (Sullivan, Mullen and Pathé, 2005, p 319), leading to unnecessary detention due to false positive findings that the individual concerned is at risk of harming others. Importantly, all of these scales are based on variables that are derived from analyzing groups, giving rise to the 'statistical truism that the mean of a distribution tells us about everyone, yet no one' (Cooke and Michie, 2010). Ian Coyle and Robert Halon (2013) have further observed that:

> The law guarantees that a decision will be made but it does not guarantee outcomes. Yet that is precisely what the law seeks to require of those engaging in the task of risk-analysis of dangerousness. It is time, once and for all, to acknowledge that estimates derived from actuarial tests cannot predict the future behavior of individuals with anything approaching that implicit in the legal minimum standards of proof.

Returning to the Loomis case, the trial court used a COMPAS report to justify incarceration. The court said that "You're identified, through the COMPAS assessment, as an individual who is at high risk to the community. In terms of weighing the various factors, I'm ruling out probation because of the seriousness of the crime and because your history, your history on supervision, and the risk assessment tools that have been utilized, suggest that your [sic] extremely high risk to re-offend." Loomis' counsel led evidence from an expert witness, Dr Thompson, who outlined the pitfalls of relying on actuarial risk assessment in a sentencing context. The State of Wisconsin did not offer any witnesses to counteract this evidence, instead arguing that the court's conclusion did not rely on the COMPAS report and, if it did, any reliance was a "harmless error".

One of the appeal points was whether the trial court's use of COMPAS at sentencing violated Loomis' constitutional right to due process because Loomis could not challenge the scientific validity of the assessment due to Northpointe's proprietary claim over the software algorithm. Loomis argued that it was unknown which criminogenic factors COMPAS utilizes, and how it weighs them. He relied on *Gardner v Florida* (430 U.S. 349, 351 (1977)). In Gardner, the defendant was convicted of first degree murder and the trial court sentenced him to death. The defendant appealed because the trial court deemed certain portions of the pre-sentence investigation report to be confidential and refused to disclose the information to counsel. The US Supreme Court held that the defendant was denied due process because the trial court had imposed a sentence "at least in part, on the basis of information which (Gardner) had no opportunity to deny or explain". The Court of Appeals accepted Loomis' submission that the same principle applied here, and concluded that the "apparent

limited ability of (the) defendants to investigate the tool" unfairly prevented them from assessing its scientific validity.

Wisconsin appealed to the US Supreme Court. In his appellate submissions, Loomis argued that:

> the only basis for COMPAS are the ipse dixit statements from Northpointe that it does what it says; that although we do not know how it weighs the criminogenic factors, we should just take the risk assessment as true. To do so, however, violates Mr. Loomis' (and other defendant's) right to due process because information upon which the trial court is relying for sentencing is secret and confidential. As the Court of Appeals noted, there is a lack of transparency.

Wisconsin, for its part, defended COMPAS (citing Brennan, 2009). Remarkably, Wisconsin argued that "Loomis claims the COMPAS report may have been inaccurate, but he cannot prove it because he does not know how COMPAS calculates risk". Quite. Instead, Wisconsin argued that Loomis knew what questions the COMPAS evaluation asked and he knew the answers to the questions – and on *that* basis he could contest the answer to specific questions on the COMPAS evaluation if he thought that the correct answer was different than the answer entered, and *that* procedure satisfied the constitutional due process requirement. Specifically, according to the State of Wisconsin, "Due process does not require disclosure of the formulas used to determine risk".

Remarkably, the US Supreme Court rejected the appeal. The Court was likely influenced by an amicus curiae brief filed, on the Court's request, by the US Solicitor-General. The Solicitor-General opined that given "the highly limited purpose for which petitioner's ability to counter the factual information on which the assessment relied, the Wisconsin Supreme Court correctly declined to find a due process violation. But that is not to say that the use of actuarial risk assessments at sentencing will always be constitutionally sound." While the issues the petition raised were conceded to be important, the Solicitor-General said that any "constitutional error in considering (the) petitioner's COMPAS score was likely (to have been) harmless".

What a remarkable occasion to apply the maxim *de minimis non curat lex*.

The Wisconsin Supreme Court has since published a guideline judgment relating to COMPAS which has rather confusingly said that while COMPAS Reports cannot be determinative, they nevertheless may be regarded as *relevant* in sentencing. *How* relevant will, it seems, remain a mystery. Perhaps tacitly acknowledging the weakness of this reasoning, the Wisconsin court has imposed several prophylactic guidelines: first, any "presentence investigation report ("PSI") containing a COMPAS risk assessment filed with the court must contain a written advisement listing" its limitations, and second, if used in sentencing, the following cautions need to be applied:

- "The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined.

- Because COMPAS risk assessment scores are based on group data, they are able to identify group of high-risk offenders-not a particular high-risk individual.
- Some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism.
- A COMPAS risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations."

On this basis, "*if used properly* with awareness of the limitations and cautions, a circuit court's consideration of a COMPAS risk assessment at sentencing does not violate a defendant's right to due process" (emphasis added).

Not long after the Supreme Court delivered its judgment rejecting the Loomis appeal, Professor Shirley Ann Jackson, President of the Rensselaer Polytechnic Institute in New York, asked the Chief Justice of the United States, the Hon John Roberts Jr., "[c]an you foresee a day when smart machines, driven with artificial intelligences, will assist with courtroom fact-finding or, more controversially even, judicial decision-making?" Roberts CJ replied, "It's a day that's here, and it's putting a significant strain on how the judiciary goes about doing things."

It seems remarkable that the constitutional right to due process would not protect the defendant in a criminal trial, and ensure that person's access to information used against them. Would Australian common law principles of procedural fairness operate to protect a person placed in a similar position in Australia?

3 "Preventive Exile" after Failing the Character Test

I'm not aware of any Australian case where a court had to consider the procedural justice implications of the use of secret algorithms. However the Australian Government is, apparently, actively considering the development of risk assessment tools, and it is conceivable that similar issues might arise. A Cabinet briefing note leaked in the first half of 2016 proposed the introduction of "a *visa risk assessment tool* that establishes an intelligence-led threat identification and *risk profiling capability* incorporating immigration as well as national security and *criminality risk* for visa applicants".[1]

Ian Coyle and I have written elsewhere (2016) about the use of character testing in the immigration system (and the next few paragraphs rely heavily on that paper). Specifically, in late 2014 the Minister for Immigration and Border Protection issued Direction 65 to supplement section 501 of the *Migration Act* 1958 (Cth), which enables the Minister or a delegate to cancel a visa held by a noncitizen convicted of an offence on the basis that they have failed a 'character test'. A person is presumed to

[1] David Lipson, 'Leaked Government document outlines tougher migration program, increased monitoring of refugees', ABC News, 4 February 2016.

fail the character test if they have a 'substantial criminal record', defined as a criminal conviction attracting a sentence (or cumulative convictions and sentences, adding up to) of 12 months or more.

The removal of people pursuant to this regime takes place without prior notice being given (presumably to prevent absconding) and is often effected in the early hours of the morning by armed personnel.[2] Once arrested and removed from their homes and families, these people are typically taken to an immigration detention centre such as Christmas Island. Christmas Island is a lovely name for an island but it is very remote – some thousands of kilometres from Australia in the middle of the Indian Ocean. It is a long way from family, friends and supports. Here, detainees may wait many months or even years for their case to be heard, assuming they are able to secure legal assistance to do so. This plainly raises significant concerns about their ability to access the justice system to challenge their removal.

With Ian Coyle, I have explored the use of a 'risk assessment' as a basis for decision-making about removals – and considered whether it is compliant with proper forensic standards. We have argued that if risk assessments are to be undertaken, they need to be undertaken properly, and with a nuanced appreciation of the limitations of forensic tools. Serious questions can be raised about the utility of actuarial risk assessment tools. However the risk remains that governments will devise regulations that remove the power of litigants and courts to question these tools. Applying knowledge of group tendencies to individual offenders within actuarial risk assessment approaches can have dire consequences when transferred to court settings in high stakes cases where liberty or citizenship is at stake (McSherry and Keyzer, 2009). Problems identified with the *use* of actuarial-based scales in relation to individual offenders were acknowledged in the guideline judgment delivered after *Loomis v Wisconsin.* But it is the gloss on *Gardner v Florida* that is significant in an age of algorithmic governance. It is difficult to prevent an involuntary shaking of the head when a company's proprietary interests are elevated above the right to liberty.

In China, a social credit system has been devised to rate people on their social and financial behavior. It is said that this new system will "allow the trustworthy to roam everywhere under heaven while making it hard for the discredited to take a single step." (Hawkins, 2017). While horrifying, and so horrifying to be scarcely believable, surveillance is not new, and has been carried out for eons. The Chinese Communist *dang'an,* or secret personal file, tracks a citizen's information from their high school grades, to their behavior at university, to their perceived political sympathies in adult life. The file can affect a person's career prospects and pension entitlements. The Tibetan writer Tsering Woeser has described the dang'an as "an invisible monster stalking you" (Jacobs, 2015).

There has been an appreciable rise in the development and deployment of risk assessment tools to judge us all. We want to remove risk from our lives and we expect our governments to do this. In the commercial sphere, we are all affected by the administrative justice meted out by electronic platforms and services that we use for

[2] *Eden v Minister for Immigration and Border Protection* [2015] FCA 780.

transport, to do shopping, and to employ assistants. We buy into these regimes by rating people ourselves.

Actuaries say that we should work with all available information, and that it would be wrong not to. But where is the place for the presumption of innocence, let alone the possibility of rehabilitation, in the coming dystopia? Do we have the legal tools to challenge the risk assessments?

References

1. Australia. Commonwealth Ombudsman, Report No 2 of 2017.
2. Tim Brennan et. al., 'Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System', (2009) 36 *Criminal Justice and Behaviour* 21.
3. David Cooke and Christine Michie, 'Limitations on Diagnostic Precision and Predictive Utility in the Individual Case: A Challenge for Forensic Practice', (2010) 34 *Law and Human Behavior* 259.
4. Ian Coyle and Robert Halon, 'Humpty Dumpty and Risk Assessment: A Reply to Slobogin', in Patrick Keyzer, ed., *Preventive Detention: Asking the Fundamental Questions*, Intersentia, 2013.
5. Ian Coyle and Patrick Keyzer, 'The removal of convicted noncitizens from Australia: is there only a 'minimal and remote' chance of getting it right?' (2016) 41(2) *Alternative Law Journal* 86
6. *Gardner v Florida* 430 U.S. 349 (1977).
7. Peter Hanks QC, 'Administrative law and welfare rights: a 40-year story from Green v Daniels to "robot debt recovery"', Australian Institute of Administrative Law Conference, Canberra, July 20, 2017.
8. Amy Hawkins, 'Chinese Citizens Wants the Government To Rank Them', Foreign Policy, May 24, 2017.
9. Andrew Jacobs, 'A Rare Look Into One's Life On File in China', Sinosphere, 15 March, 2015.
10. Patrick Keyzer, 'The International Human Rights Parameters for the Preventive Detention of Serious Sex Offenders', in Bernadette McSherry and Patrick Keyzer, eds., *Dangerous People: Policy, Prediction and Practice*, Routledge, 2011.
11. Bernadette McSherry and Patrick Keyzer, *Sex Offenders and Preventive Detention*, The Federation Press, 2009.
12. Melissa Perry and Alexander Smith, 'iDecide: the Legal Implications of Automated Decision-Making', University of Cambridge, Cambridge Centre of Public Law Conference, 2014
13. Danny Sullivan, Paul Mullen and Michele Pathe, 'Legislation in Victoria on Sexual Offenders: Issues for Health Professionals', (2005) 183(6) *Medical Journal of Australia* 318
14. Mariano-Florentino Cuellar, 'Artificial Intelligence and the Administrative State', The Regulatory Review: A Publication of the Penn Program on Regulation, December 19, 2016.